

Pruning a Minimum Spanning Tree

Leonidas Sandoval Junior

Insper, Instituto de Ensino e Pesquisa

September 6, 2011

Abstract

This work employs some techniques in order to filter random noise from the information provided by minimum spanning trees obtained from the correlation matrices of international stock market indices prior to and during times of crisis. The first technique establishes a threshold above which connections are considered affected by noise, based on the study of random networks with the same probability density distribution of the original data. The second technique is to judge the strength of a connection by its survival rate, which is the amount of time a connection between two stock market indices endure. The idea is that true connections will survive for longer periods of time, and that random connections will not. That information is then combined with the information obtained from the first technique in order to create a smaller network, where most of the connections are either strong or enduring in time.

1 Introduction

Minimum spanning trees are networks of nodes that are all connected by at least one edge so that the sum of the edges is minimum, and which present no loops. This kind of tree is particularly useful for representing complex networks, filtering the information about the correlations between all nodes and presenting it in a planar graph. Because of this simplicity, minimum spanning trees have been widely used to represent some important financial structures, namely the structures of stock exchanges [1]-[22], of currency exchange rates [23]-[28], of world trade [29]-[32], commodities [33], of GDPs (Gross Domestic Products) [34], corporations [35]-[36] and, most important for this article, of the world financial markets [9], [18], [37]-[40].

Like any other representation of real world interactions, it is subject to a great amount of random noise, which is sometimes difficult to isolate. This work employs some techniques in order to filter random noise from the information provided by a minimum spanning tree. The networks that are represented are obtained from the correlations between international stock exchange indices prior to and during years of well known international financial crises, namely the 1987 Black Monday, the 1997 Asian Financial Crisis, the 1998 Russian Crisis, the crisis after September, 11, 2001, and the Subprime Mortgage crisis of 2008. The reason for choosing international stock exchange indices is because one has some idea of which indices should be more correlated based on information that comes from other sources that are not the correlation matrix, such as geographical proximity, cultural affinity, and strength of commercial relations [29], and also because one can then follow the interactions of stock markets at different periods of time and in different volatility regimes.

One of the techniques is to establish a threshold above which connections are considered affected by noise based on the study of random networks with the same probability density distribution of the original data, what is obtained by shuffling the original data so that correlations between the various nodes are basically random. This eliminates both random connections and true ones, what can be shown by using arguments of geographical, cultural, and financial links between the stock exchanges. This effect is particularly strong for emerging or small financial markets.

The second technique is to judge the strength of a connection by its survival rate, which is the amount of time a connection between two stock market indices endure. The idea is that true connections will survive for longer periods of time, and that random connections will not. That information is then combined with the information coming from the first technique in order to create a smaller network, where most of the connections

are either strong or enduring in time. The resulting network is a k -minimum spanning tree [41]-[43], with some additional nodes and edges attached to it, and the result is then studied using the tools of network analysis.

This study brings out important information, like the existence of an American and of an European cluster, both strongly connected in themselves but weakly connected with one another. The study also brings out the importance of the Netherlands in the 80's and 90's, and the emergence of France as the main hub for the European cluster. The study also shows how a Pacific Asian cluster was formed during the last two decades, and of how smaller, weakly interacting clusters, like the one formed by some members of the former Yugoslavia, or the one of Arab indices, have formed in the last decade.

2 Correlations and distance

The time series of financial market indices encode an enormous amount of information about the way they relate to each other. Part of this information may be captured by the correlation matrix of their log-returns, which are defined as

$$S_t = \ln(P_t) - \ln(P_{t-1}) \approx \frac{P_t - P_{t-1}}{P_t}, \quad (1)$$

where P_t is the value of a certain index at time t and P_{t-1} is the value of the same index at time $t - 1$.

There are many measures of correlation between elements of time series, the most popular being the Pearson correlation coefficient. The drawback of this correlation measure is that it only detects linear relationships between two variables. Two different measures that can detect nonlinear relations are the Spearman and the Kendall tau rank correlations, which measure the extent to which the variation of one variable affects other variable, without that relation necessarily being linear. In this work, we choose Spearman's rank correlation, for it is fairly fast to calculate.

The correlation matrix between log-returns of financial market indices may be used to classify the markets from which it is made into clusters, or into "neighbors". For that, we use the process first developed in [1] and used in most articles, which considers a suitable distance function between the markets based on the correlation function, and then uses the Minimum Spanning Tree technique in order to establish links between them.

An Euclidean metric must fulfill the axioms $d_{ij} = 0 \Leftrightarrow i = j$, $d_{ij} = d_{ji}$, and $d_{ij} \leq d_{ik} + d_{kj}$. In this work, I shall use a different metric from reference [1], which is a nonlinear mapping of the Pearson correlation coefficients between stock returns. The metric to be considered here differs from the aforementioned metric because it is a linear realization of the Spearman rank correlation coefficient between the indices that are being studied:

$$d_{ij} = 1 - c_{ij}, \quad (2)$$

where c_{ij} are elements of the correlation matrix calculated using Spearman's rank correlation. This distance goes from the minimum value 0 (correlation 1) to the maximum value 2 (correlation -1). The linear mapping is used in order to best identify the relation between correlation and distance.

Using the metric given by (2), one obtains a distance matrix, over which can be applied the Minimum Spanning Tree technique, which consists of choosing one of the indices (nodes) and finding the next node which is closest to that one, and then to link them. One then considers the cluster formed by those two nodes and finds the next node that is closest to any of the members of that cluster, and links it to the node it is closest to inside the cluster. The process then goes on until there are no free nodes left.

As an example, using the correlation matrix for 1987, starting from the S&P 500 index, the Nasdaq is the closest index to it. So, one establishes a link between them and looks for the next node (index) that is closest to one of the two indices in the cluster. It turns out to be the S&P TSX (Canada), which is closest to the node Nasdaq. So, one establishes a link between Nasdaq and S&P TSX. Following this procedure, one obtains a planar graph with no loops that shows relationships between the markets that are being considered.

Although the minimum spanning tree is a valuable tool for the study of connections between different assets or indices, it has some limitations. One of them is that it sometimes overstates connections that are dim, leading sometimes to the illusion that two indices are strongly connected when they are not. Some countries that have long distances from any other country and are actually wanderers in the landscape of the correlation ties are connected almost at random in a minimum spanning tree.

One way to decide whether a connection made by the Minimum Spanning Tree technique is random or not is to compare the results of real data with those obtained for a minimum spanning tree that is built from random interactions. This is discussed in the next section.

3 Simulations with random matrices

Figure 1 shows the minimum spanning tree obtained from the Spearman correlation of 40 randomly generated returns, based on a Gaussian distribution with zero mean and standard deviation 2. One characteristic of this graph is that distances vary little, averaging 0.8. Figure 2 displays the frequency distribution of distances for this particular minimum spanning tree.

The other remarkable fact is the distribution of node degrees, where a node degree is the number of connections a particular node has with other nodes (as an example, node 14 has three connections, so it has node degree 3). Figure 3 shows the frequency distribution for this particular minimum spanning tree realization of the random data.

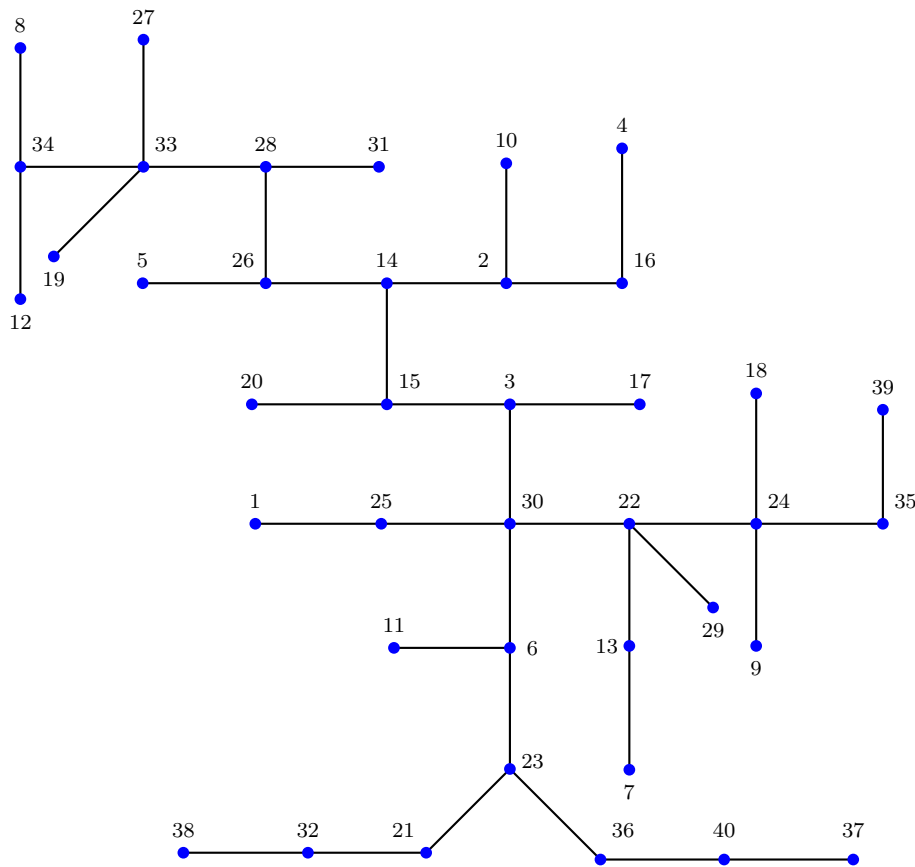


Figure 1: minimum spanning tree for randomly generated data with forty vertices.

In general, minimum spanning trees generated from real data will differ significantly from this behavior: the distances vary over a larger scope than the distances calculated for random data, and the node degree frequency distribution is also remarkably different, with nodes connecting to many more nodes than expected for random connections. Nevertheless, some of the distances obtained from real data will be the result of random noise, and comparison with the results obtained for random data shall be useful in order to separate what are real connections from random ones.

Actually, we shall not use minimum spanning trees based on random Gaussian returns, for the distributions of returns of financial assets are rarely Gaussian. Instead, we shall follow the approach already used by many authors [refs] and work with flushed data, which is obtained by reordering the time series of each index at random so as to break down any correlation between indices but preserve the same probability distribution.

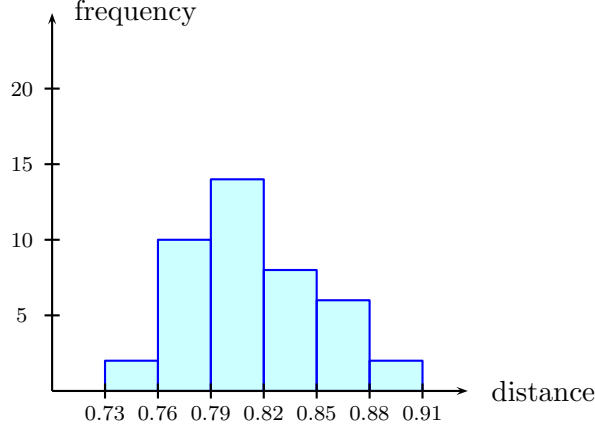


Figure 2: frequency distribution of the distances for a minimum spanning tree obtained from 40 random returns.

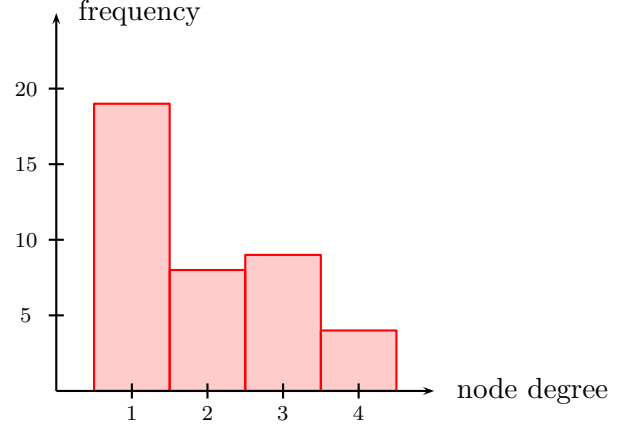


Figure 3: frequency distribution of the node degrees for a minimum spanning tree obtained from 40 random returns.

4 Minimum spanning trees

In what follows, we shall calculate the minimum spanning trees for a diversity of international stock exchange indices around key periods in the past decades. The data were chosen so as to present a variety of economies and also offer a diversity of volatility regimes. More specifically, we chose the years of the last major world financial crises, namely the 1987 Black Monday, the 1997 Asian Financial Crisis, the 1998 Russian Crisis, the crisis after September, 11, 2001, and the Subprime Mortgage crisis of 2008, and the years that preceded them (with the exception of the Asian Financial Crisis). As those crises usually occurred in the second and sometimes in the first semester of a year, we separated data into two semesters for each year that was studied, so that we may then study the evolution of the interactions among markets previous to and during financial crises when seen through the eyes of minimum spanning trees. The data used in this article is essentially the same as in [44], and details of the indices used, and their codes, can be found in that article (or in Appendix A of the present article), which also offers a comprehensive list of bibliography on the use of Random Matrix Theory in Finance.

4.1 1987 - Black Monday

We first analyze the first and second semesters of 1986 and 1987. The crisis that occurred in October, 1987, caused the loss of trillions of dollars worldwide, as some markets lost some 30% of their values in a few days. We use, for this study, the indices of 16 markets, more specifically, those from the USA (S&P 500 and Nasdaq), Canada, Brazil, UK, Germany, Austria, Netherlands, India, Sri Lanka, Japan, Hong Kong, Taiwan, South Korea, Malaysia, and Indonesia. The names of the indices and stock exchanges to which they belong are listed in Appendix A. The data comprise small as well as very large markets, and offer a variety of nations and cultures across three continents.

Each minimum spanning tree is drawn with the indices represented by 2 to 4 letters (also listed in Appendix A) that make it easier to recognize the country each index belongs to. The vertices are also colored according to geographical region: orange for North America, green for South America, blue for Europe, brown for central Asia, and red for Pacific Asia. For each minimum spanning tree, we generated flushed data based on the same data used for calculating the minimum spanning tree. Distances that are within the region predicted for randomized data are represented in dashed lines, the other distances being represented by solid lines. All distances shown in the graphics are in scale, and the same scale is preserved throughout the article, so indices with strong connections appear much closer than ones which have low correlation, and the many minimum spanning trees can be easily compared.

Figure 4 shows the minimum spanning tree obtained for the first semester of 1986. For this interval of time, the average randomized distance is 0.81 ± 0.01 (mean plus or minus the standard deviation), the result

of 1000 simulations, with average minimum 0.77 ± 0.04 and average maximum 0.92 ± 0.02 . So, we represent distances below 0.77 as solid lines and, above this threshold, as dashed lines. Only the connections between S&P 500 and Nasdaq, Nasdaq and Canada, Nasdaq and the UK, Germany and Netherlands, can be trusted. Such connections like Canada and Sri Lanka, or Netherlands and Hong Kong, should be ignored, but other connections, which seem true ones, like between Japan, South Korea, and Malaysia, are also considered as probably random if the same standards are used. It is possible that distances between indices are inside the interval considered as random noise but are, nevertheless, true connections. Our method of establishing a threshold and ignoring any distance above it may eliminate some of those true connections.

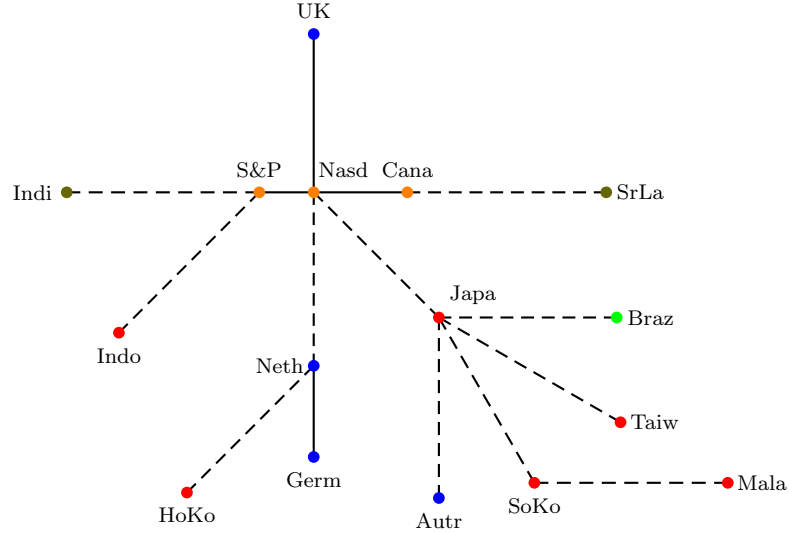


Figure 4: minimum spanning tree for the first semester of 1986.

Figure 5 does the same for data regarding the second semester of 1986. Solid lines represent distances below 0.77 ± 0.04 (the result of yet more 1000 simulations) and dashed lines represent distances above that. Again, the lines connecting S&P, Nasdaq, Canada, and the UK are solid, so it is a cluster that kept its stability over the two timespans that have been studied until now. Other solid connection is again between Germany and the Netherlands, and one new solid connection appears between Canada and the Netherlands. From the dashed lines of figure 4, only the connection between S&P and Indonesia remains, what speaks against the stability of the weaker connections of the minimum spanning trees.

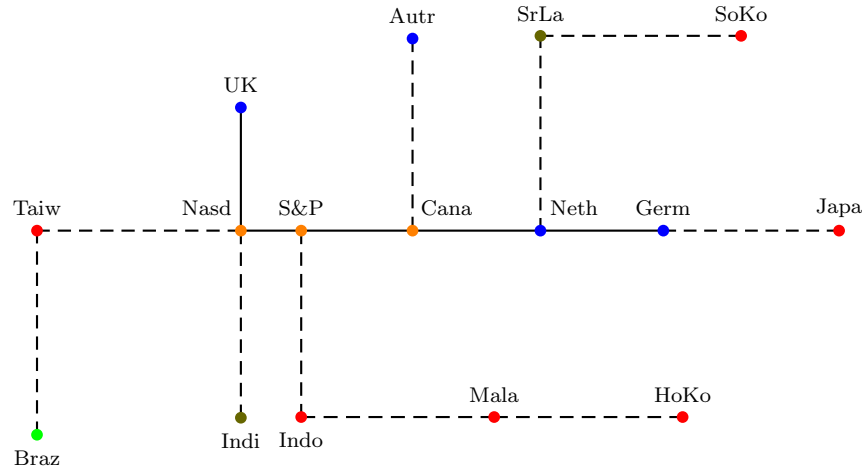


Figure 5: minimum spanning tree for the second semester of 1986.

Figures 6 and 7 show the minimum spanning trees obtained from data concerning 23 indices, from the USA (S&P and Nasdaq), Canada, Brazil, the UK, Ireland, Germany, Austria, Netherlands, Sweden, Finland, Spain, Greece, India, Sri Lanka, Bangladesh, Japan, Hong Kong, Taiwan, South Korea, Malaysia, Indonesia,

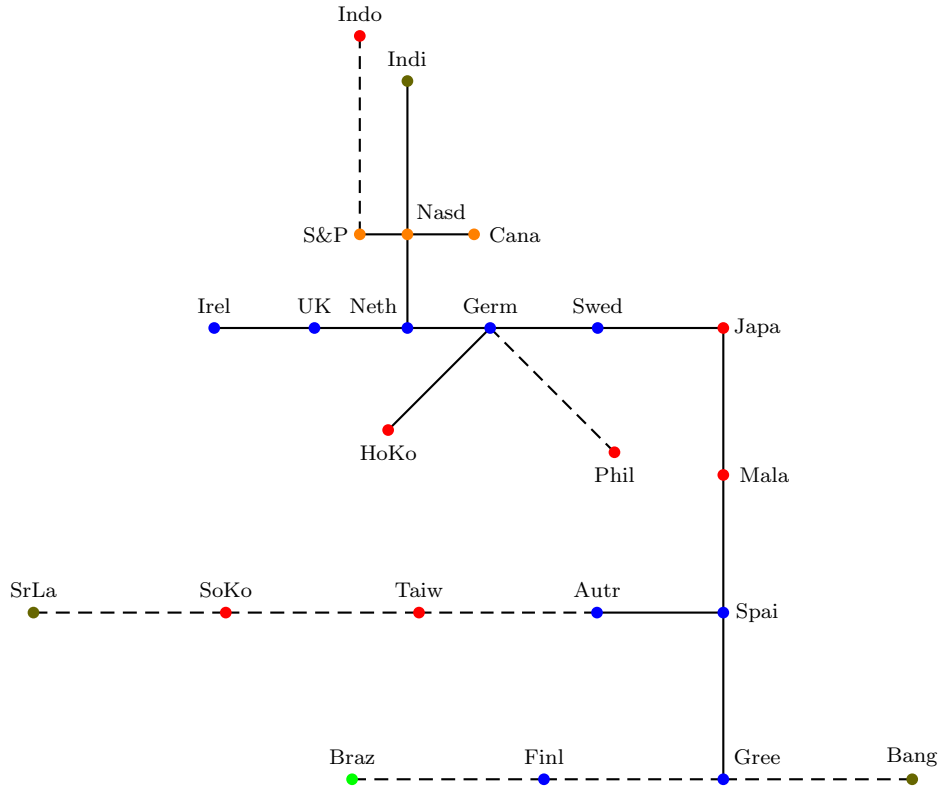


Figure 7: minimum spanning tree for the second semester of 1987.

4.2 1997 and 1998 - Asian Financial Crisis and Russian Crisis

Now, we jump in time to the next crisis: the Asian Financial Crisis, which started with the devaluation of the Thai baht, the currency of Thailand, and spread to most of the Pacific Asian countries. This led to the Russian crisis, which began with the economic crisis in Russia due to the fall of its exports, and also because of the war in Chechnya, and spread to many more countries. Now we have data from 57 indices for 1997 - S&P, Nasdaq, Canada, and Mexico (North America, represented in orange), Costa Rica, Bermuda, and Jamaica (Central America and the Caribbean, represented in darkgreen), Brazil, Argentina, Chile, Venezuela, and Peru (South America, in green), UK, Ireland, France, Germany, Switzerland, Austria, Belgium, Netherlands, Sweden, Denmark, Finland, Norway, Iceland, Spain, Portugal, Greece, Czech Republic, Slovakia, Hungary, Poland, and Estonia (Europe, in blue), Turkey (Eurasia, in lightblue), Israel, Lebanon, Saudi Arabia, Oman, Pakistan, India, Sri Lanka, and Bangladesh (West and Central Asia, in brown), Japan, Hong Kong, China, Taiwan, South Korea, Thailand, Malaysia, Indonesia, and Philippines (Pacific Asia, in red), Australia (Oceania, in black), Morocco, Ghana, Kenya, South Africa, and Mauritius (Africa, in magenta). For 1997, we have no data about Russia, which is added to the data concerning 1998 (which then has 58 indices). All indices and symbols used for them are best explained in Appendix A.

Figures 8 and 9 show clear clusters, one formed by European indices, and the other formed by American ones. For the first semester of 1997, Netherlands is the main hub for the European cluster, and there is a subdivision of the American cluster between North and South American indices. Australia and South Africa are also clearly connected with the European cluster, as is the UK, which tended to belong with the American cluster before. Canada connects both clusters in both periods of time. Malaysia and Indonesia are connected by a solid connection just under the threshold. We also have a solid connection between Hong Kong and Austria and another between Iceland and Mauritius, both apparently random in nature. The threshold for the first semester of 1997 is 0.70 ± 0.03 .

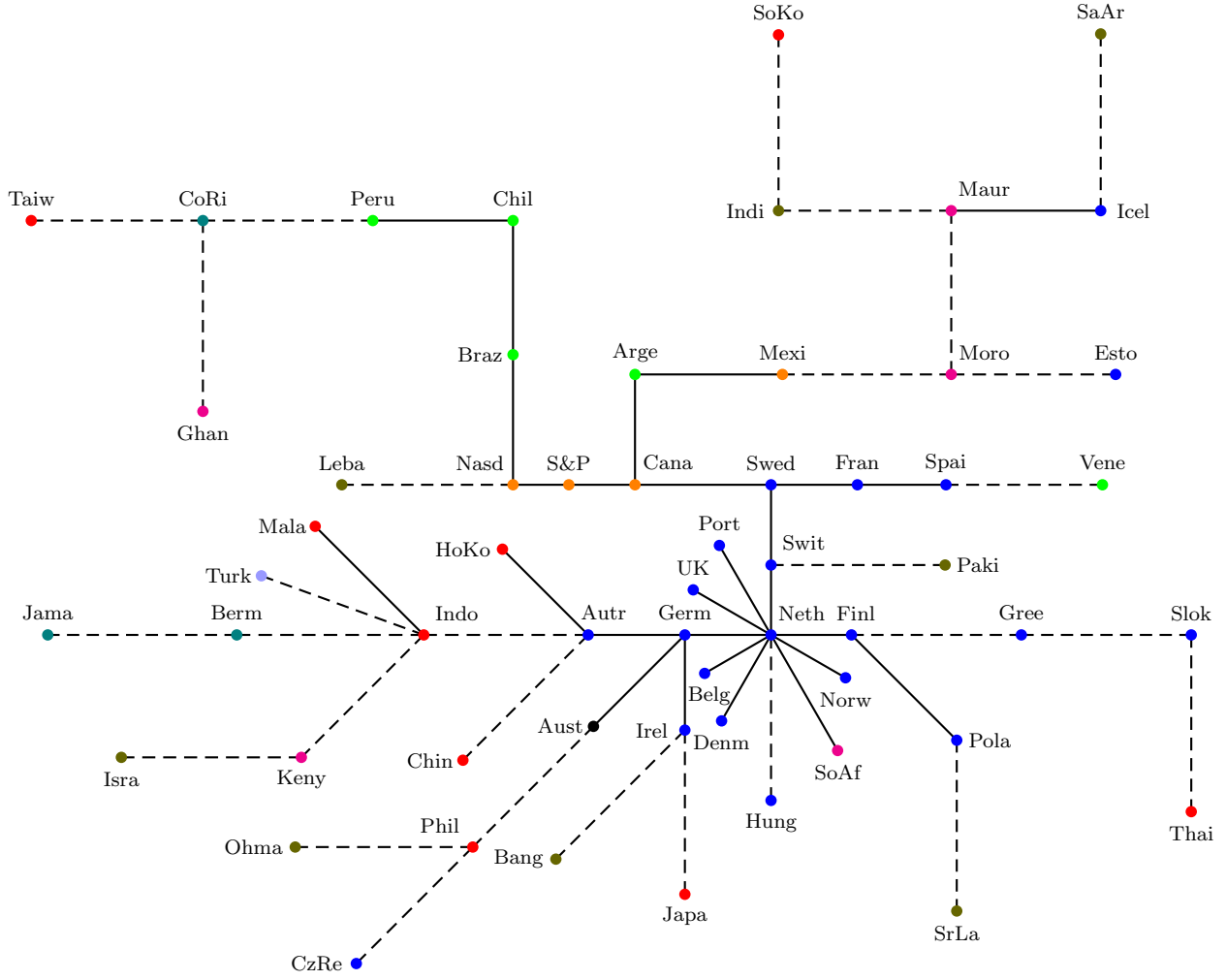


Figure 8: minimum spanning tree for the first semester of 1997.

For the second semester of 1997, structures are very similar, but with Germany taking over itself some of the centrality of Netherlands. Some other differences are Hungary and Israel connecting with Europe, Hong Kong connecting now with Germany, Japan connecting with Australia, and the formation of a Pacific Asian cluster comprised of Hong Kong, Malaysia, Indonesia, and Philippines. The threshold now is 0.71 ± 0.03 . Some connections, like Peru and Austria, seem to be random.

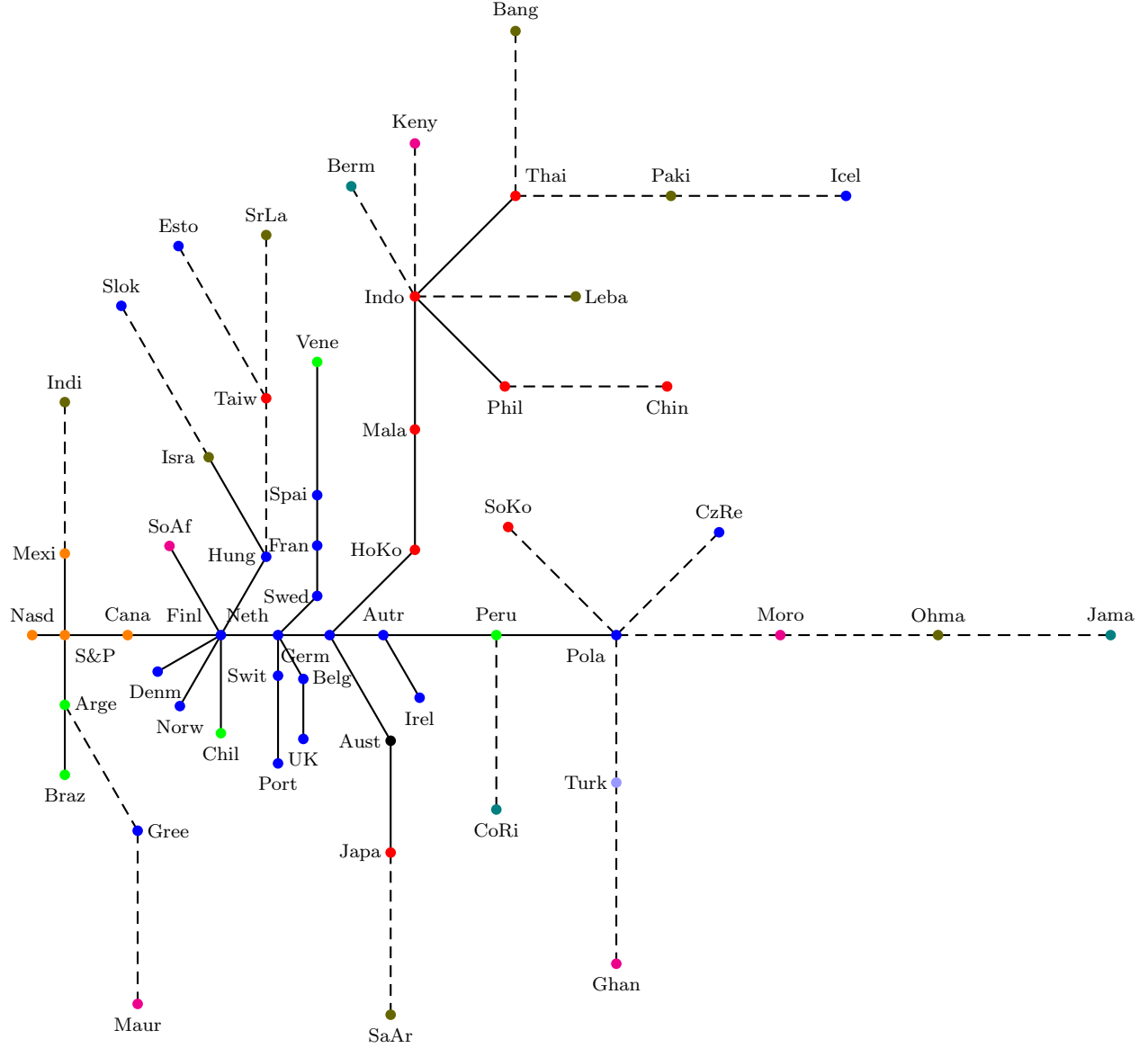
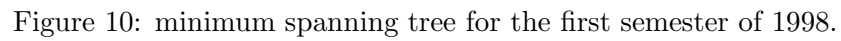


Figure 9: minimum spanning tree for the second semester of 1997.

For 1998, we add Russia to the indices that were used in 1997. Figures 10 and 11 show the minimum spanning trees for both, respectively, the first and the second semesters of that year. The threshold for the first semester of 1998 is 0.70 ± 0.03 , what leaves below it the usual American and European clusters, a Pacific Asian cluster, and a new cluster centered around Russia (which is colored light blue for Eurasia), comprised of Hungary, the Czech Republic, Estonia, and Poland. Strange connections are those between Greece and Turkey with Canada, Indonesia-Chile-Switzerland, Venezuela-Norway, and Australia, South Korea and Pakistan with Poland. All connections above the threshold seem to be random.

The threshold for the second semester of 1998 is 0.71 ± 0.03 . The resulting graph (figure 10) shows a similar behavior to the first semester of the same year, except that Russia and Turkey are now more connected with Europe, and that now the Pacific Asian cluster is split into three parts, the main part being the cluster comprised of Japan, Hong Kong, Philippines, and Australia, although Australia belongs to Oceania. Chile is now connected with the American cluster via Mexico. Also to be noticed is that clusters clearly shrunk in the second semester of 1998, in particular the European cluster, whose indices agglomerated around the Netherlands.



4.3 2001 - Burst of the Dot-Com Bubble and September 11

The year 2001 saw two crises: the first one was internal, the result of excessive speculation on equities of internet-based companies, known as dot-com; the second, was purely exogenous, the result of the largest terrorist attack in history, perpetrated against the USA. Here, we analyze the years 2000 and 2001, in order to check how the two crises influenced the minimum spanning trees resulting from data relative to many stock market exchanges in the world.

We now have 74 indices: S&P, Nasdaq, Canada, and Mexico (North America, in orange), Panama, Costa Rica, Bermuda, and Jamaica (Central America and the Caribbean, in darkgreen), Brazil, Argentina, Chile, Venezuela, and Peru (South America, in green), UK, Ireland, France, Germany, Switzerland, Austria, Italy, Malta, Belgium, Netherlands, Luxembourg, Sweden, Denmark, Finland, Norway, Iceland, Spain, Portugal, Greece, Czech Republic, Slovakia, Hungary, Poland, Romania, Estonia, Latvia, Lithuania, and Ukraine (Europe, in blue), Russia and Turkey (Eurasia, in lightblue), Israel, Palestine, Lebanon, Jordan, Saudi Arabia, Qatar, Oman, Pakistan, India, Sri Lanka, and Bangladesh (Western and Central Asia), Japan, Hong Kong, China, Mongolia, Taiwan, South Korea, Thailand, Malaysia, Singapore, Indonesia, and Philippines (Pacific Asia, in red), Australia (Oceania, in black), Morocco, Tunisia, Egypt, Ghana, Nigeria, Kenya, South Africa, and Mauritius (Africa, in magenta). Figures 12 and 13 show the minimum spanning trees for the first and second semesters of 2000.

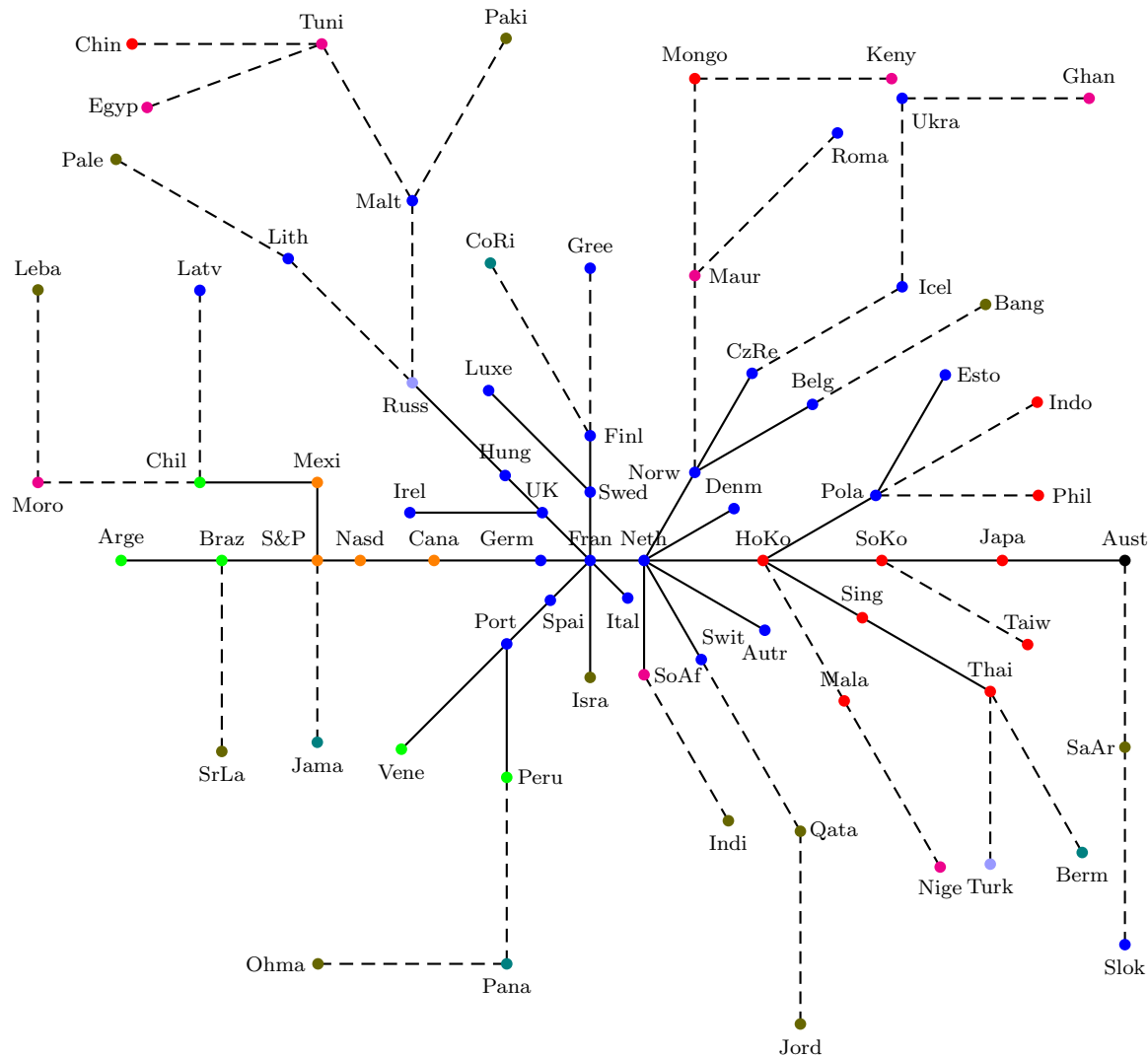


Figure 12: minimum spanning tree for the first semester of 2000.

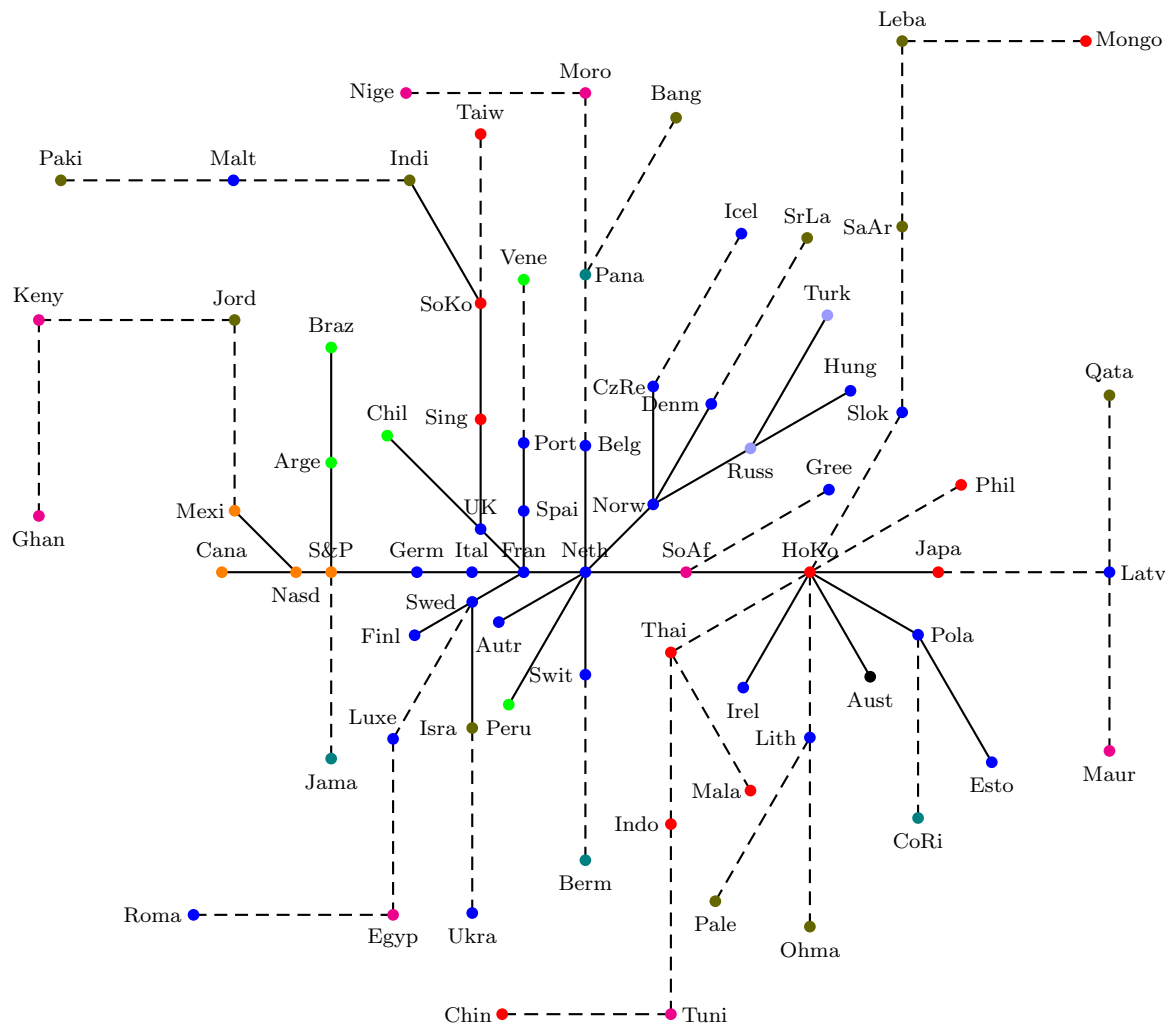


Figure 13: minimum spanning tree for the second semester of 2000.

Starting from the first semester of 2000, depicted in figure 12, we establish a threshold 0.70 ± 0.03 . The American cluster is almost unmodified, and the European cluster has now two main hubs: France and Netherlands. Israel and South Africa are still connected with the European cluster, and Australia is now definitely part of a Pacific Asian cluster, comprised of itself, Japan, Hong Kong, South Korea, Singapore, and Taiwan. Above the threshold, there is one connection that seems significant, the one between Hong Kong and Malaysia, and many other connections that seem random in nature.

Figure 13 shows the minimum spanning tree for the second semester of 2000, with the threshold 0.69 ± 0.03 . What one can see from it is that the American and the European clusters are not very altered, but the Pacific Asian cluster has been again pulverized into some looser connections. If we consider some of the connections above the threshold, then Hong Kong becomes a hub for some Pacific Asian indices and also for some European ones.

For 2001, we now have 78 indices, for we added Bulgaria, Kazakhstan, Vietnam, New Zealand, and Botswana to the indices that were already being used. The minimum spanning tree for the first semester of 2001 is depicted in figure 14 with threshold 0.69 ± 0.03 . The South American cluster is detached from the North American one, and France slowly assumes the role of main hub for European indices, replacing Netherlands. We have some semblance of a cluster of Pacific Asian indices, together with those from Oceania (Australia and New Zealand). Other structures below the threshold remain, basically, unaltered.

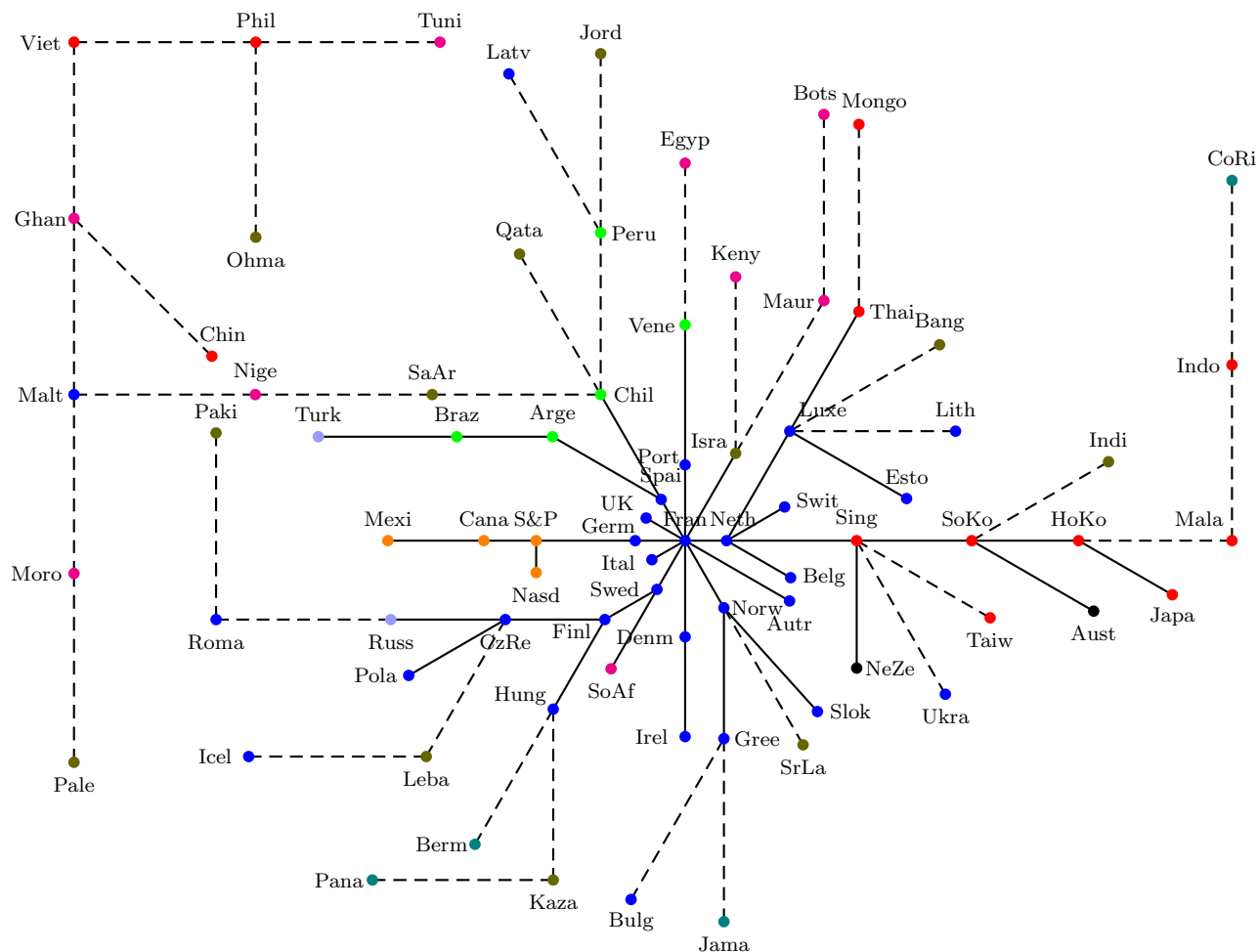


Figure 14: minimum spanning tree for the first semester of 2001.

For the second semester of 2001 (figure 15), the threshold is 0.69 ± 1 . France is now the undisputable hub for Europe, and both the American and the Pacific Asian clusters are clearly defined. The newcomer index, New Zealand, connects with Australia, as it would be expected, but a good number of indices are still loosely connected, and the connections shown in dashed lines are probably all due to random noise. The exception are the loose connections between some Arab indices (Jordan-Qatar-Ohman-Egypt).

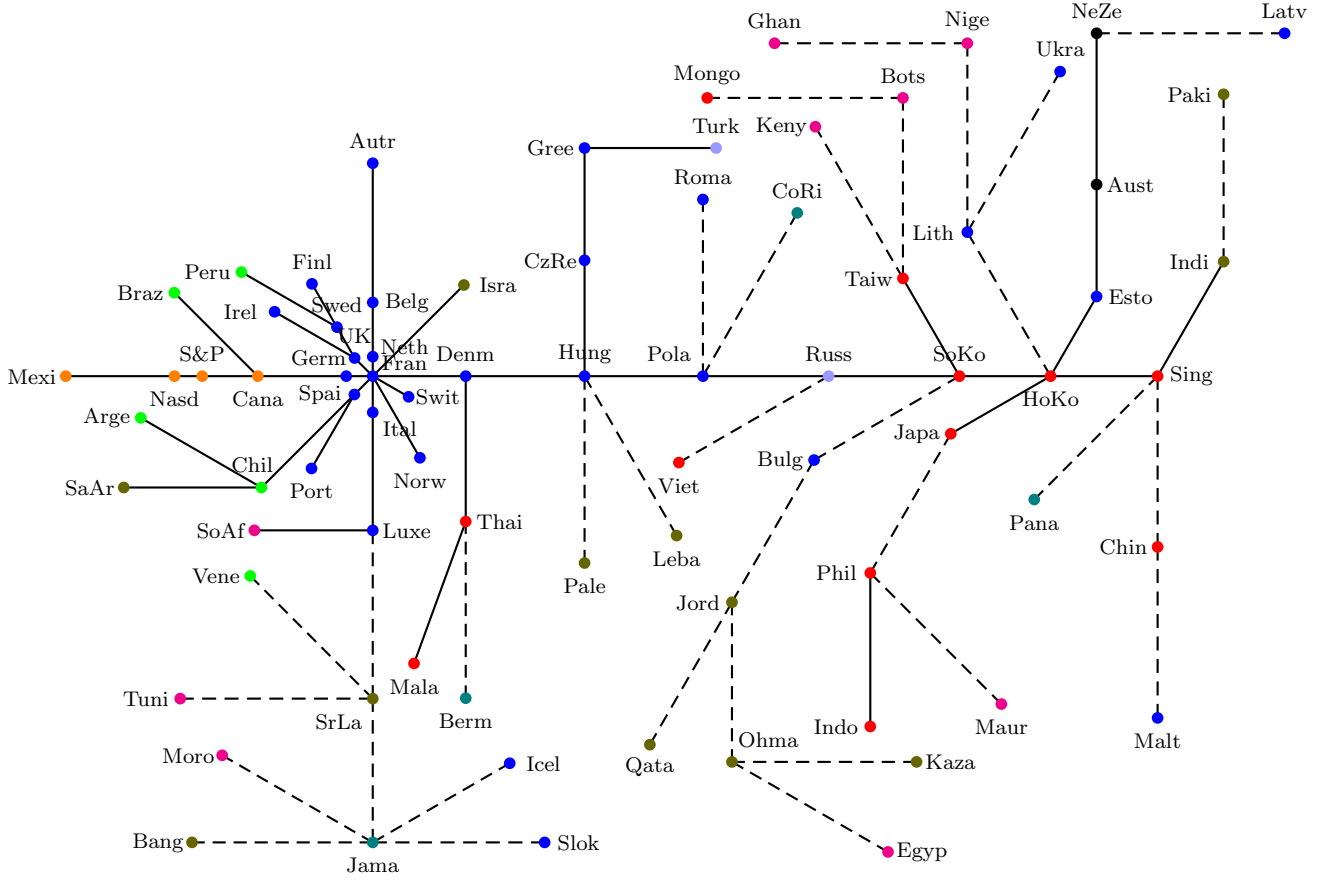


Figure 15: minimum spanning tree for the second semester of 2001.

4.4 2008 - Subprime Mortgage Crisis

The last crisis to be analyzed is the so called Subprime Mortgage Crisis, which had its origins in the USA and then spread to the world as a Credit Crisis. It started at the end of 1997 and reached its peak in 2008, with the bankruptcy of major investment banks, and has its effects up to now. We here analyze the years 2007 and 2008, using 92 indices, namely S&P and Nasdaq from the USA, and indices from Canada, Mexico, Panama, Costa Rica, Bermuda, Jamaica, Brazil, Argentina, Chile, Colombia, Venezuela, Peru, UK, Ireland, France, Germany, Switzerland, Austria, Italy, Malta, Belgium, Netherlands, Luxembourg, Sweden, Denmark, Finland, Norway, Iceland, Spain, Portugal, Greece, Czech Republic, Slovakia, Hungary, Serbia, Croatia, Slovenia, Bosnia and Herzegovina, Montenegro, Macedonia, Poland, Romania, Bulgaria, Estonia, Latvia, Lithuania, Ukraine, Russia, Kazakhstan, Turkey, Cyprus, Israel, Palestine, Lebanon, Jordan, Saudi Arabia, Kuwait, Bahrain, Qatar, United Arab Emirates, Oman, Pakistan, India, Sri Lanka, Bangladesh, Japan, Hong Kong, China, Mongolia, Taiwan, South Korea, Thailand, Vietnam, Malaysia, Singapore, Indonesia, Philippines, Australia, New Zealand, Morocco, Tunisia, Egypt, Ghana, Nigeria, Kenya, Tanzania, Namibia, Botswana, South Africa, and Mauritius. The number of indices grew due to the emergence of new financial markets and also of new countries, like those that became independent after the end of the former Yugoslavia (although Yugoslavia began disintegrating in 1991, it took some time for the new countries to develop solid stock exchanges).

We start from figure 16, which shows the minimum spanning tree for the first semester of 2007 with threshold 0.68 ± 0.03 . Again, one can see a large agglomeration of indices around France, and a very compact European cluster of indices. The American cluster remains stable, now with the addition of Colombia. The presence of a Pacific Asian cluster (together with the indices from Oceania) is now much clearer. Namibia is tightly connected with South Africa, what is to be expected, for Namibia has been under the economic and political influence of South Africa for the past six decades. Cyprus is connected with Greece, what is to be expected given their common cultures and ethnicities (Cyprus' stock exchange is located at the Greek part of the island). There is also the appearance of a small cluster of some of the former members of the extinct Yugoslavia: Croatia, Bosnia and Herzegovina, and Macedonia. Among the weaker links, one can devise an Arab cluster, comprised of Bahrain, Qatar, Kuwait, United Arab Emirates, Lebanon, and Saudi Arabia. These connections, although weak, do not seem random. One may also notice that Singapore is the main hub for Pacific Asian indices, probably due to the country's expertise in finance.

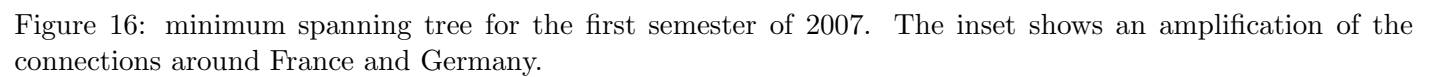


Figure 17 shows the minimum spanning tree for the second semester of 2007, also with threshold 0.68 ± 0.03 . Notice that the graph is more compact, with most of the European indices agglomerating around France. The basic American cluster and the Pacific Asian cluster are still solid, although smaller economies, like Colombia, Peru, Vietnam, and Mongolia are not attached to their geographical clusters. Namibia and South Africa keep strongly connected, and South Africa and Israel keep strongly connected with Europe. Some links which were previously weak are now reinforced, like the ones between the former members of Yugoslavia, and the links between some Arab countries. One can also see two pairs of African indices, Tanzania-Ghana and Nigeria-Tunisia that may or may not be the result of random noise.

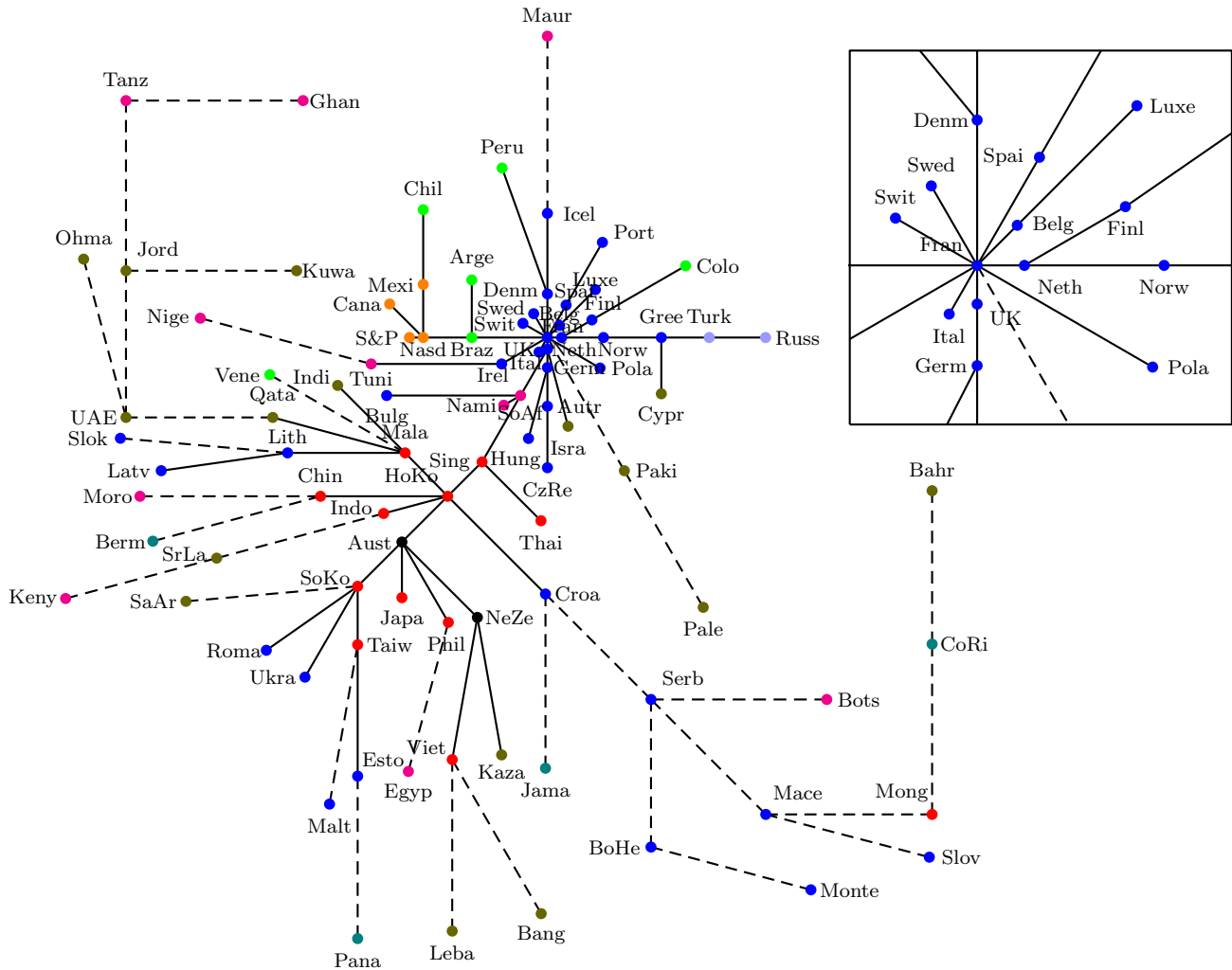


Figure 17: minimum spanning tree for the second semester of 2007. The inset shows an amplification of the connections closest to France.

Figure 18 shows the minimum spanning tree for the first semester of 2008, with threshold 0.68 ± 0.03 . Here, again, one can clearly see the presence of the American, European, Pacific Asian, and Arab clusters. Connections between African indices, with the exception of the pair Namibia - South Africa, haven't lasted. Some of the former members of Yugoslavia, Croatia and Macedonia, seem now more integrated with Europe, and a weakly connected cluster is made of Serbia, Montenegro, and Bosnia and Herzegovina. We also have a strong connection between Ukraine and Russia.

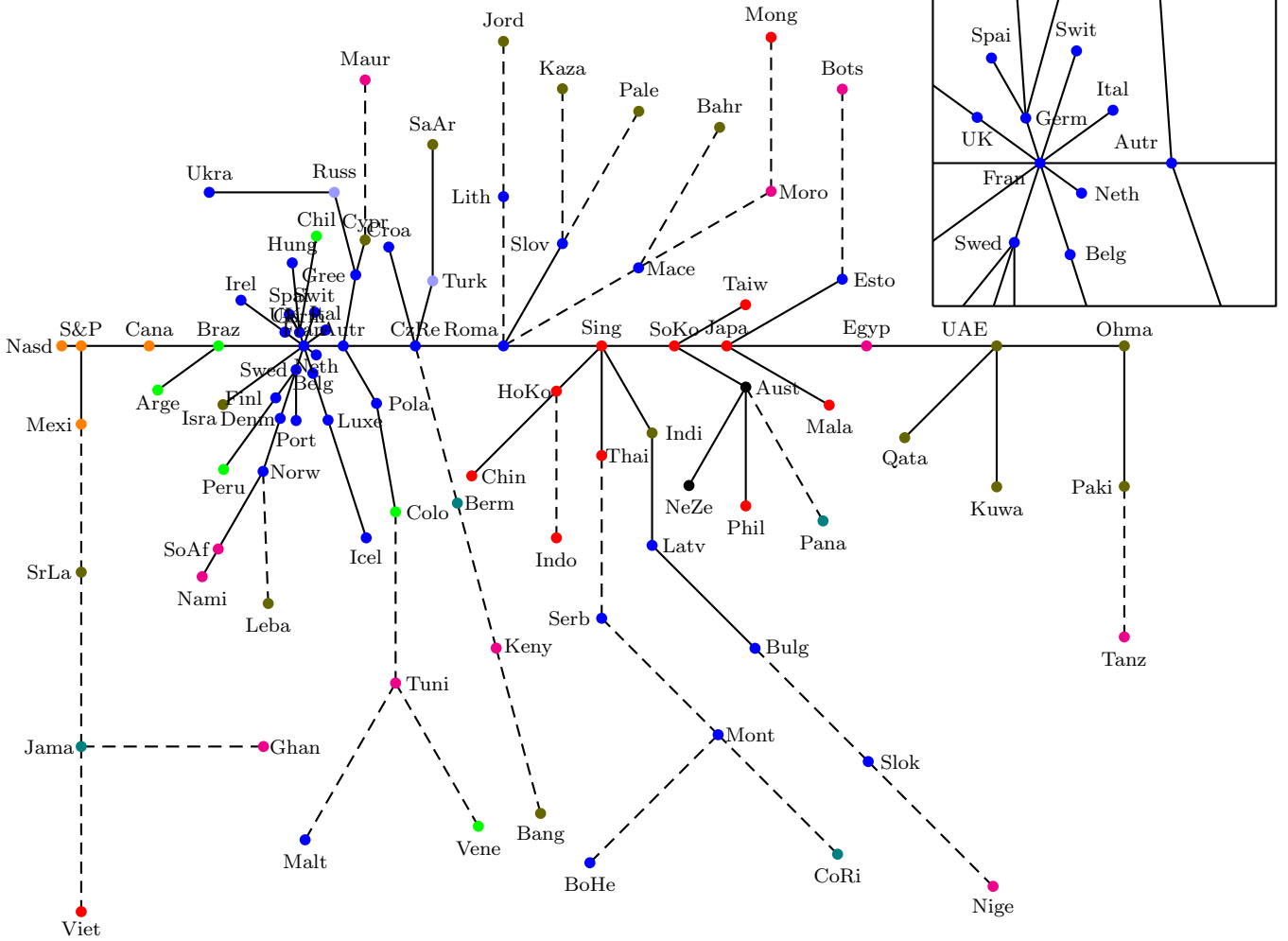


Figure 18: minimum spanning tree for the first semester of 2008. The inset shows an amplification of the connections closest to France.

The second semester of 2008 is represented in figure 19 for a threshold 0.68 ± 0.03 . This is a high volatility period, especially after the main crisis, which occurred in September of that same year. The network has shrunk due to the increase in correlations between markets during and after the peak of the crisis: S&P and Nasdaq have nearly joined, and France also nearly merges with other European indices. The Czech Republic now works as a hub for some Eastern European indices, and Australia and Singapore are the main hubs for the Pacific Asian cluster. There is a clear cluster of other Eastern European indices, namely Slovakia, Serbia, Bulgaria, Estonia, Lithuania, and Latvia, and a well formed cluster of Arab indices (United Arab Emirates, Qatar, Palestine, Jordan, Oman, Bahrain, and Kuwait). Saudi Arabia is still isolated, though. There are also connections between Botswana and Ghana, and between Egypt, Morocco, and Tanzania.

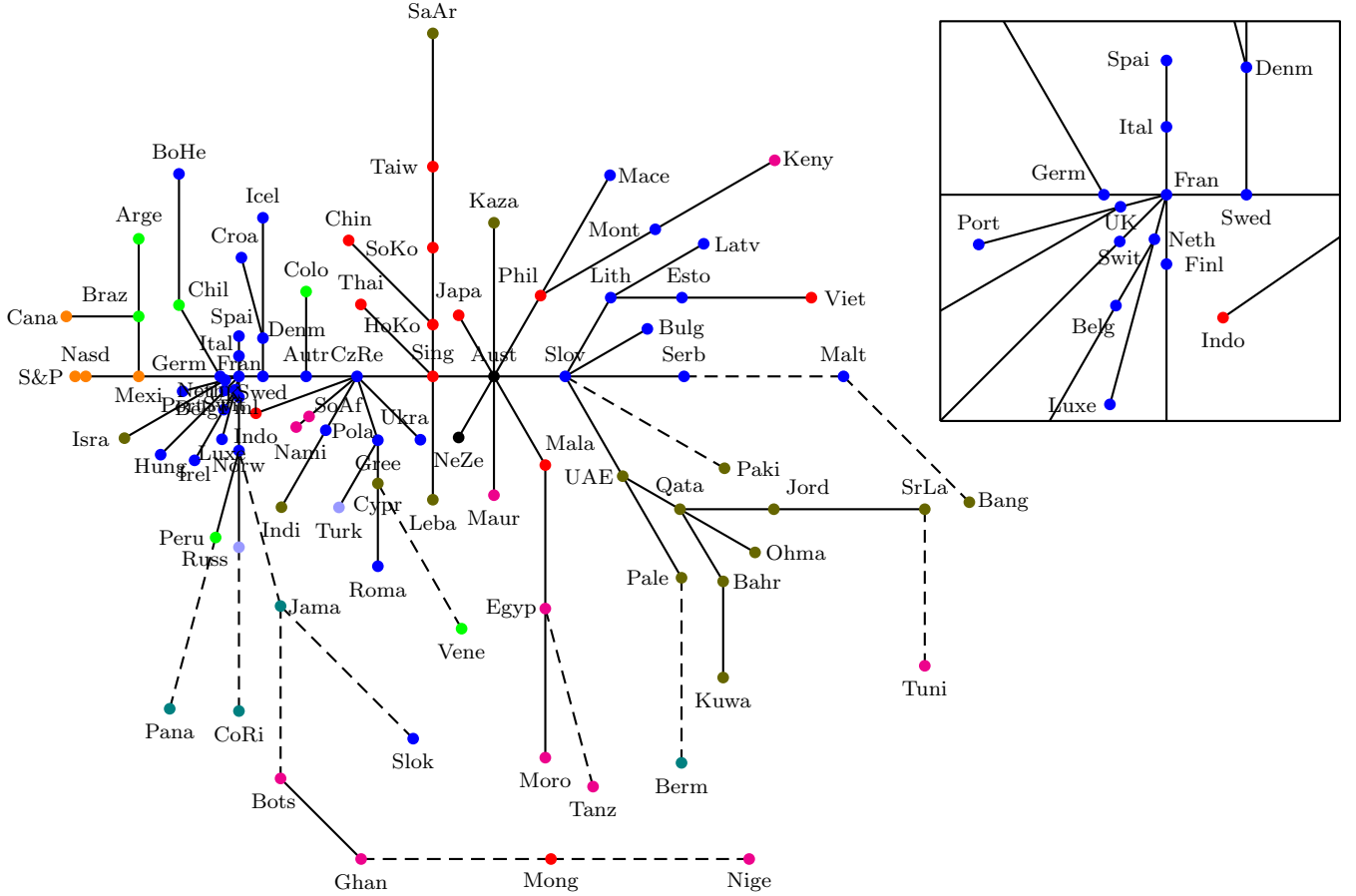


Figure 19: minimum spanning tree for the second semester of 2008. The inset shows an amplification of the connections closest to France.

5 Pruning

As we could see from the previous section, the connections between indices in a minimum spanning tree, which in our case is based on distances which are themselves based on the correlations between them, is plagued by random noise, which may place suspicion on the weakly connected indices. By performing simulations on shuffled data, which preserve the frequency distributions of the times series used but destroys correlations between them, we could see that, in average, distances above a threshold of approximately 0.7 are typical of random interactions. So, any connection above this threshold may be due to random noise. When those connections are removed, the resulting network is called a k-minimum spanning tree, which is applied to the study of some networks [41]–[43]. Nevertheless, some weaker connections may fall beyond the threshold, and using the threshold as a cut-off will probably eliminate some real and important connections from our results.

One way to separate true connections from false ones is to hypothesize that if two indices are truly connected, then that connection must be enduring in time. How long the connection must endure is open to debate, but

the longer a connection lasts, the more reliable it is. So, in this section we develop a method for testing the survivability of connections in a minimum spanning tree and use it to determine which connections are resilient in time and which are not.

Robustness of connections of a network have already been studied using stock returns in [45]-[46], and pruning was used in [47] based on the survival ratio of connections [48]-[49], although the sort of pruning is different from the one adopted in the present article.

We begin by considering an array, called the *adjacency matrix* A , whose elements A_{ij} are 1 when there is a connection between indices i and j and zero otherwise. We shall consider each element of this matrix as a function of time and consider sequences $A(t_0), A(t_1), \dots, A(t_n)$ of snapshots of such matrix at different times. We consider matrices calculated over periods of 60 days in a window that moves one day at a time. Such a choice offers a compromise between eliminating some statistical noise for calculating correlation matrices using a small amount of data and not extending our results over a long period of time, when the structure of the global market may change by itself. We shall use the year 2008 as an example.

Our first attempt at testing survivability follows the approach of other researchers [45]-[49], which is to define a survival matrix R as the result of the product of a sequence of adjacency matrices in time. So, we may define the elements of such a survival matrix as

$$R_{ij} = \prod_{k=1}^n A_{ij}(t_k) , \quad (3)$$

where n is the number of sliding windows used.

If the connection is strong, then it shall be present at all times, but if it is zero in a single period of time, then the entire product will be zero as well. Applying this method to the data concerning 2008, then we obtain the following results: since a minimum spanning tree of n nodes has $n - 1$ connections, then we start with 91 connections in the first calculation, done over the first 60 days of that year; moving the window five days, 53 of those connections survive; for 20 days, the connections that survive are 18; for forty days, we have 11 connections that survive, and so on. Table 1 shows the results obtained, and the graph in figure 20 shows the evolution of surviving connections in time (solid line). Also in table 1 and figure 20 are the average number of surviving connections in time resulting from 1000 simulations with randomized data based on the same time series as the real data (dashed line). Dotted lines represent the average simulated values plus or minus their standard deviation.

Timespan (days)	Surviving connections	Surviving random connections
0	91	91
5	53	40.5
10	38	25.0
20	18	10.9
40	11	2.0
60	9	0.3
80	8	0.0
100	8	0.0
120	8	0
140	5	0
160	3	0
180	3	0
190	3	0

Table 1: surviving connections for matrix A over periods of time.

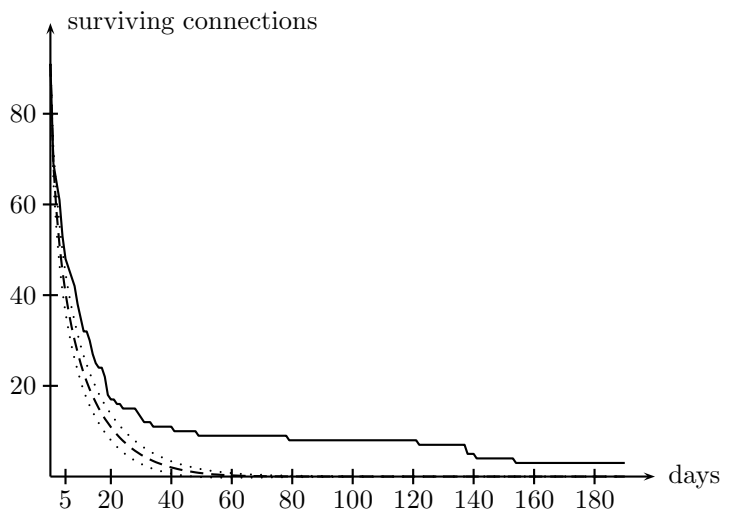


Figure 20: surviving connections as a function of time for A .

As one can see, the number of connections decreases exponentially in time, a consequence of the fact that one single zero value eliminates the survival rate of a connection. The three connections that survive after 160

days are the ones between S&P and Nasdaq, Greece and Cyprus, and Namibia and South Africa. These are all reasonable connections, and their survival in time is not a surprise. The five connections that survive up to 140 days are, besides those already cited, the ones between Belgium and Luxembourg, and between Japan and South Korea. The eight connections that survive after 120 days include the connections between France and Netherlands, Japan and Singapore, and between Taiwan and South Korea. For 60 days of survival, we also have a connection between Austria and the Czech Republic, and for 40 days of survival, connections between Germany and Spain, and between Latvia and India, this last one indicating that probably connections made by random noise may survive for some days, although for randomized data, obtained by shuffling the log returns for 2008, the survivability falls much sharper than with real data, and survivability for over 50 days is virtually zero.

Even though this seems to be a good method for pruning a minimum spanning tree, it is a little too radical, for it eliminates any connections that are disrupted for a brief period of time, but that overall are stable in time. An example would be some connections among European indices, or among members of the Arab cluster. One index may be connected to one element of its cluster and at a later time connect with a different member of the same cluster. This does not indicate that the connections of that index with the others of its own cluster is inexistent, as the multiplication of the coefficients of the adjacency matrix would lead us to believe, so an alternative method should be devised in order to try to distinguish enduring connections from those that are feeble.

One such alternative is to also consider connections of the second order. A node i may be connected with another node j via a third node k . Such connections can be computed if we consider the square of the adjacency matrix A . By taking $A + A^2$, one is considering all first and second order connections between indices, so that an index may be connected to another at certain times, and then be connected to another index that is connected with the previous one. We may then define a new survival matrix by

$$S_{ij} = \prod_{k=1}^n A_{ij}^2(t_k) , \quad (4)$$

where n is the number of sliding windows used and A_{ij}^2 are the elements of matrix $A + A^2$.

Starting from 152 connections in the first calculation, done over the first 60 days of that year, and moving the window five days, 85 of the original connections survive; sliding the window for 20 days, the connections that survive are 26; for forty days, 14 connections that survive, going down to three surviving connections after 160 days. Table 2 shows the results obtained, and the graph in figure 21 shows the evolution of surviving connections in time (solid line). Table 2 and figure 22 also exhibit the results of the average values obtained from 1000 simulations with randomized data (dashed line for the average and dotted lines for mean plus or minus standard deviation).

Timespan (days)	Surviving connections	Surviving random connections
0	152	152
5	85	52.3
10	55	30.0
20	26	12.1
40	14	2.2
60	10	0.3
80	9	0.0
100	9	0.0
120	9	0.0
140	6	0
160	3	0
180	3	0
190	3	0

Table 2: surviving connections for $A + A^2$ over periods of time.

So, now we have a measure of how many connections (first and second order) survive through time. The problem is, we are measuring the survival of connections just for the first days of the year, so every connection is being judged by its behavior in a specific period of time. In order to fix that, one may measure the survivability of connections starting from different times during the year, or during a semester, since we are calculating the minimum spanning trees based on a half-yearly data. Then, one may sum over all the resulting measures of survivability and obtain a number that states the percentage of times that a certain connection exists between two indices.

As an example, for the first semester of 2008, using a moving window of 60 days, the connections (first and second order) between S&P and Nasdaq survive 100% of the time. The connections (first and second order) between S&P and Canada survive 49% of the time, and those between the S&P and Panama do not survive at all for five days in any period of time. We may hypothesise that there is a direct relation between survivability and the strength of the connection (first order), given by one minus the distance calculated for the minimum spanning tree. The graph in figure 22 shows that this hypothesis is wrong: there are several connections that endure in time yet are not considered in the minimum spanning tree and many connections in the minimum spanning tree do not endure in time. The same type of graphic is obtained if one includes all correlations between the indices, and not just the ones relative to the connections that appear in the minimum spanning tree graph.

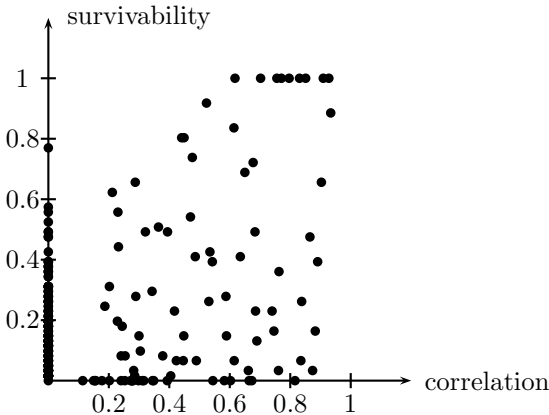


Figure 22: survivability of connections as a function of their Spearman rank correlation.

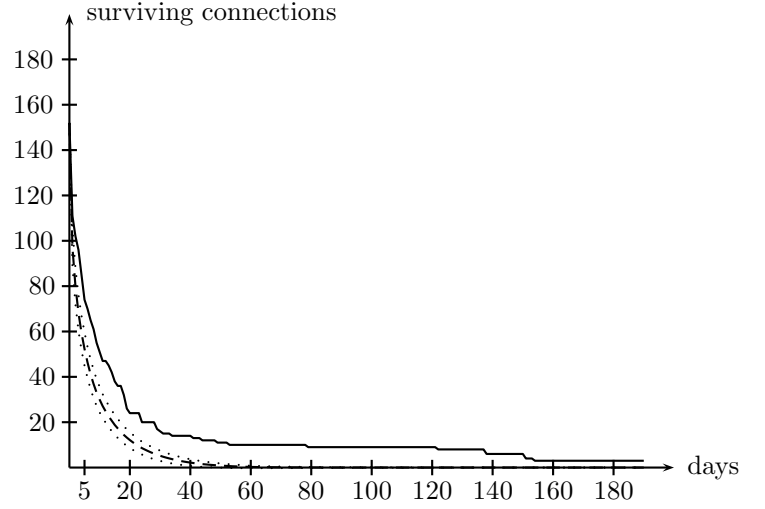


Figure 21: surviving connections as a function of time for $A + A^2$.

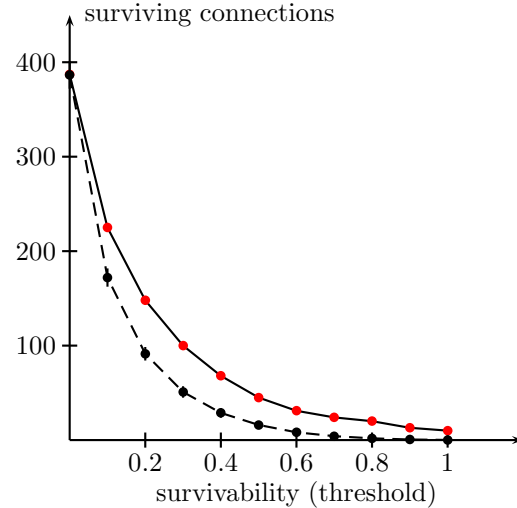


Figure 23: surviving connections as a function of minimum survivability.

Two examples help shed some light into this relationship (or lack of relationship) between survivability and correlation. One of the dots of the graph in figure 22 corresponds to the connection Germany-Italy, which has survivability 77% but does not appear in the minimum spanning tree. The connection France-Italy, which has survivability 3%, appears in the minimum spanning tree. The reason for that is that the correlation between France and Italy is slightly larger, 0.874 against 0.868, than the correlation between Germany and Italy. So, when the minimum spanning tree is being built, the link France-Italy is preferable to the link Germany-Italy. Another interesting example is the link Costa Rica - Montenegro, which has survivability 62% but correlation 0.21.

So, survivability and correlation do not go hand in hand: there are strong connections that do not survive long and weak connections that survive for longer periods of time, although connections that are stronger tend to survive longer. We must now decide if survivability may be used in order to reduce the size of a minimum spanning tree without damaging links that are strong or enduring in time, and at the same time maintaining the bi-dimensional, non-intersecting characteristics that make the minimum spanning tree such a valuable representation of the information in a network.

Figure 23 shows the number of surviving connections of the minimum spanning tree built for the first semester of 2008 as a function of a survivability threshold (solid line). The first dot corresponds to the number of all connections of first and second orders in the minimum spanning tree (387 connections), the second dot corresponds to the number of connections that survive at least in 10% of the windows considered for at least five days, and so on.

The only connections with 100% survivability and which exist in the minimum spanning tree are S&P - Nasdaq, France - Netherlands, Austria - Czech Republic, Belgium - Luxembourg, Greece - Cyprus, Japan - South Korea, Hong Kong - Singapore, Taiwan - South Korea, and Namibia - South Africa, all of them strong connections. There is no connection that is weak and that survives 100% of the time.

On the same graph of figure 23, we plot (dashed line) the survivability of connections of the average of 100 simulations with randomized data (the error bars are displayed in the figure, but are too small to be seen). One can see that the survivability of connections obtained from randomized data decreases more rapidly than the survivability obtained from the real data. One may use this graph in order to decide which is the minimum survivability requirement for a connection to be considered real. The proximity of the two curves does not make it an easy task.

Some light may be shed if one studies the survivability of the connections that are above the distance threshold obtained by considering randomized data. For the first semester of 2008, that threshold was 0.68, indicating that distances above this value in the minimum spanning tree have a stronger possibility of being due to noise. Above this threshold, the connections that survive the most are those between Slovenia and Palestine, and between Costa Rica and Montenegro, which survive 60% of the time, both of them being unlikely connections if one considers other factors others than the correlation between the indices. Only strong connections (distances below 0.68) survive more than 70% of the time. For the data concerning other years, weak correlations may survive more than 80% of the time.

We shall then consider the following rule: if a connection is below the distance threshold established by using randomized data, then it shall be considered strong and real; if the connection is above the distance threshold but survives more than 80% of the time being considered in the computation of the minimum spanning tree, then it shall also be considered real, for it is unlikely that a random connection would survive to such extent; otherwise, the connection is considered as due to random noise and removed from the graph. This is obviously a somewhat biased procedure, although based on the comparison of results between real and randomized data, and by careful analysis of the types of connections that are made below that threshold, and it is the main result we are extracting from our calculations so far.

Similar behavior occurs for other data, sampled from different time intervals. By establishing the same rules for defining which connections are strong and which are not, we then obtain the following pruned minimum spanning trees (figures 24 to 39). The only weak connections that survive the pruning procedure are those of Hong Kong - Malaysia for the second semester of 1986, Nasdaq-Netherlands and Netherlands-Japan for the first semester of 1987, Chile-Morocco and Malta-Tunisia for the first semester of 2000, and China-Tunisia for the second semester of 2000.

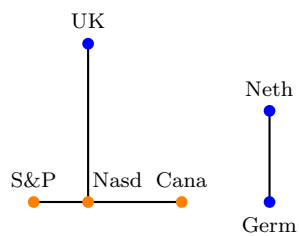


Figure 24: pruned minimum spanning tree for the first semester of 1986.

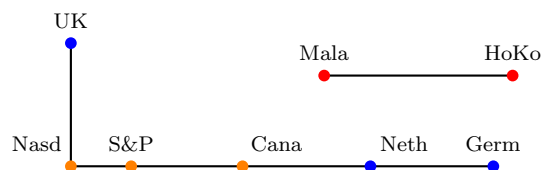


Figure 25: pruned minimum spanning tree for the second semester of 1986.

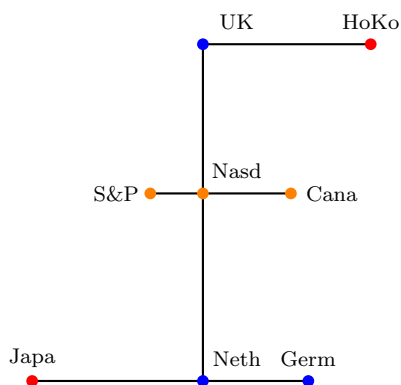


Figure 26: pruned minimum spanning tree for the first semester of 1987.

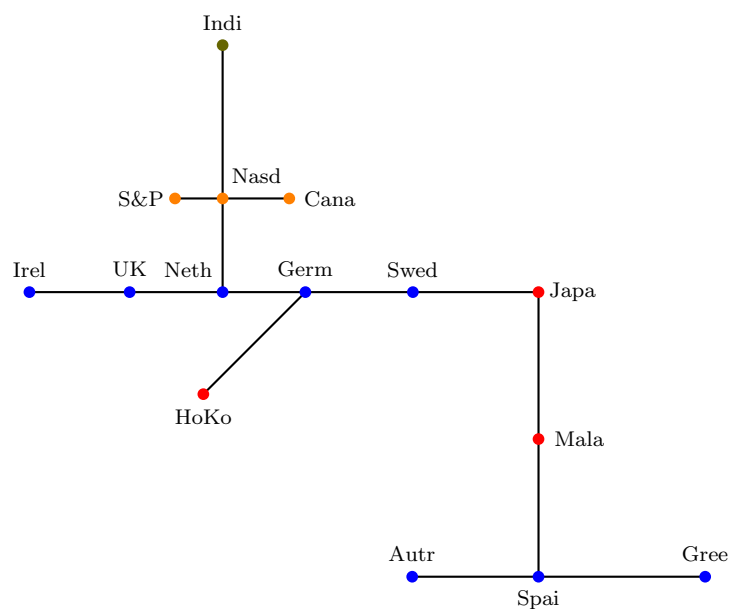


Figure 27: pruned minimum spanning tree for the second semester of 1987.

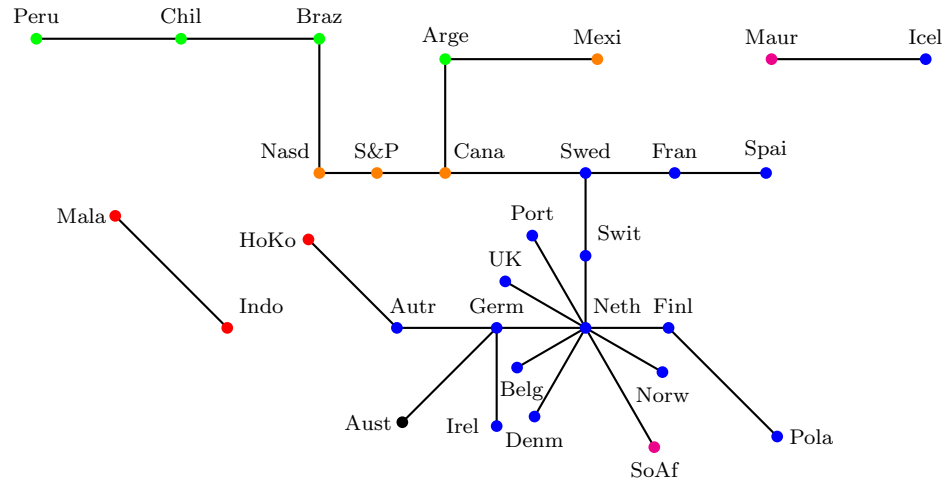


Figure 28: pruned minimum spanning tree for the first semester of 1997.

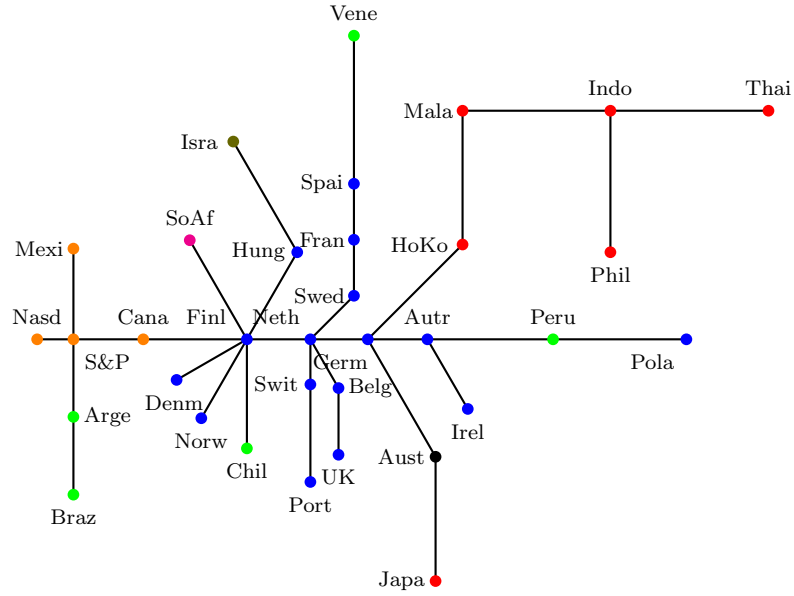
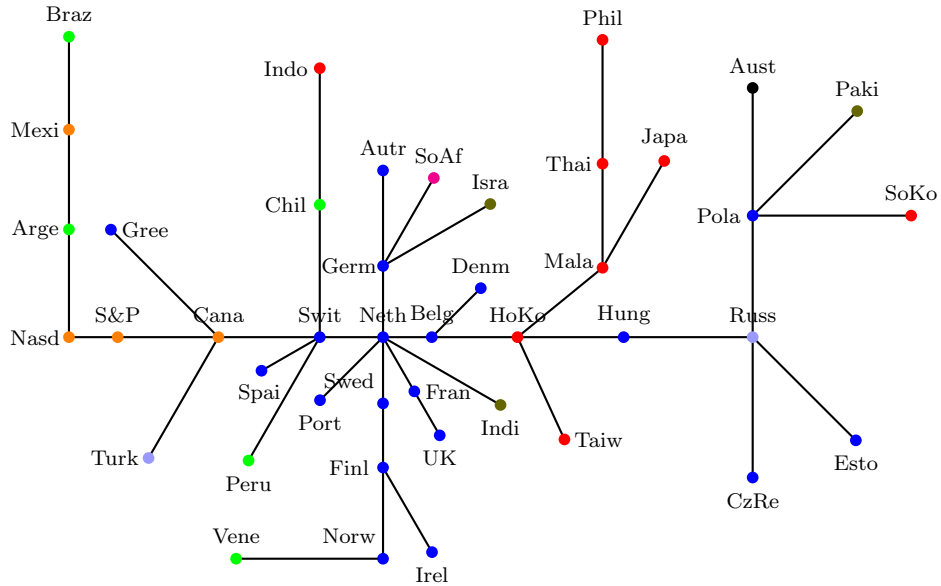


Figure 29: pruned minimum spanning tree for the second semester of 1997.



The graph displays 36 countries as nodes, connected by lines representing relationships. The nodes are color-coded by region: Europe (blue), Asia (red), Africa (pink), Latin America (orange), and Oceania (green). The graph shows a central cluster of European countries with various other countries connected to them. The layout is a force-directed graph where nodes are positioned based on their relationships.

26

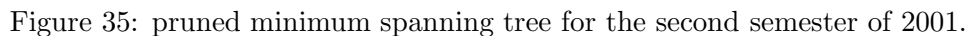


Figure 36: pruned minimum spanning tree for the first semester of 2007. The inset shows an amplification of the connections around France and Germany.

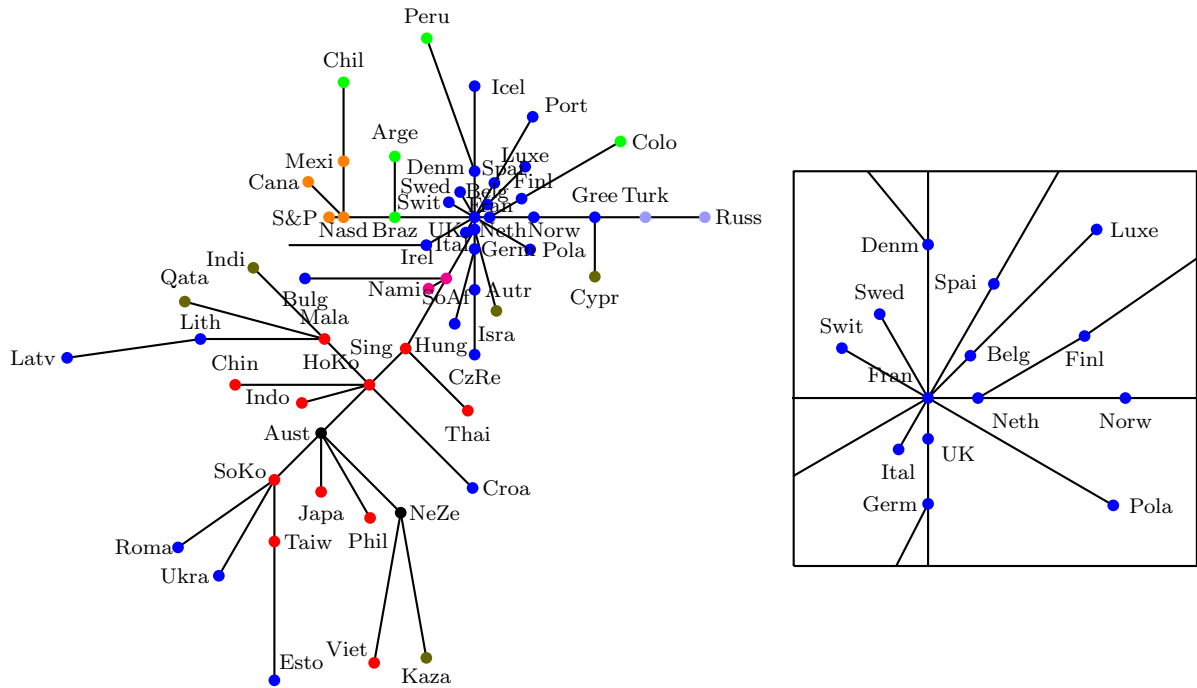


Figure 37: pruned minimum spanning tree for the second semester of 2007. The inset shows an amplification of the connections closest to France.

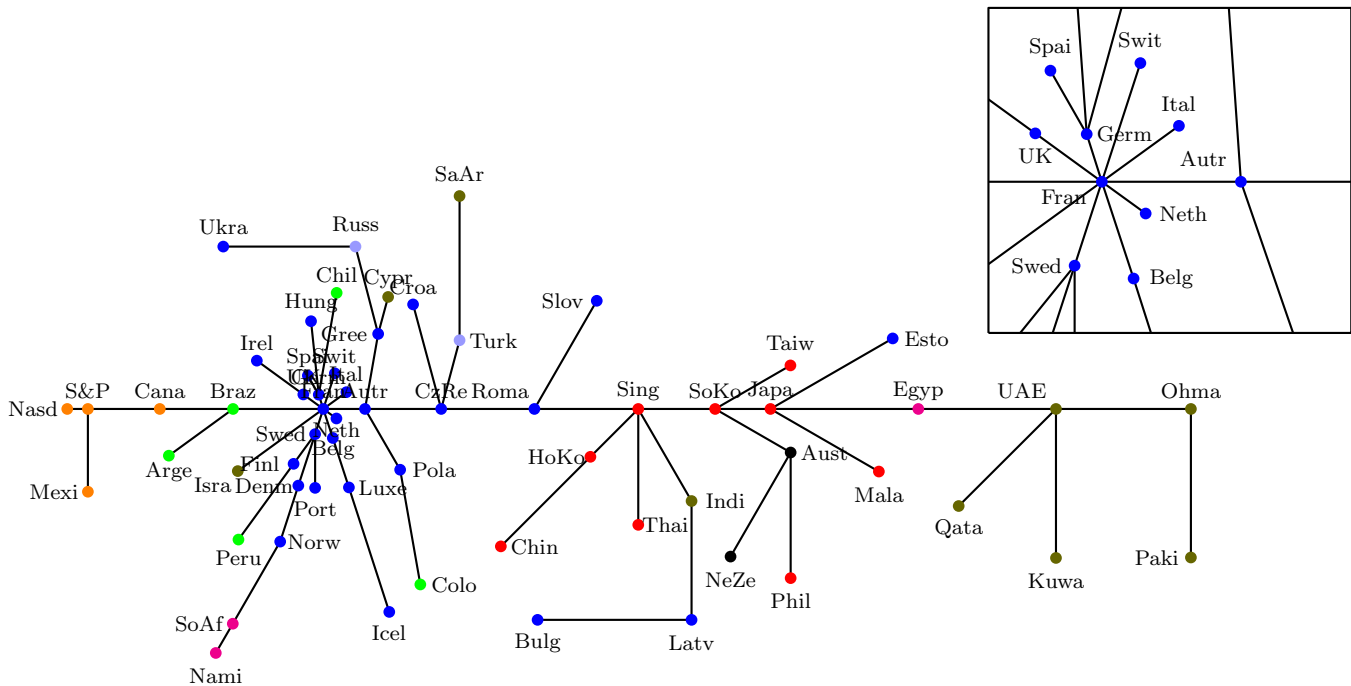


Figure 38: pruned minimum spanning tree for the first semester of 2008. The inset shows an amplification of the connections closest to France.

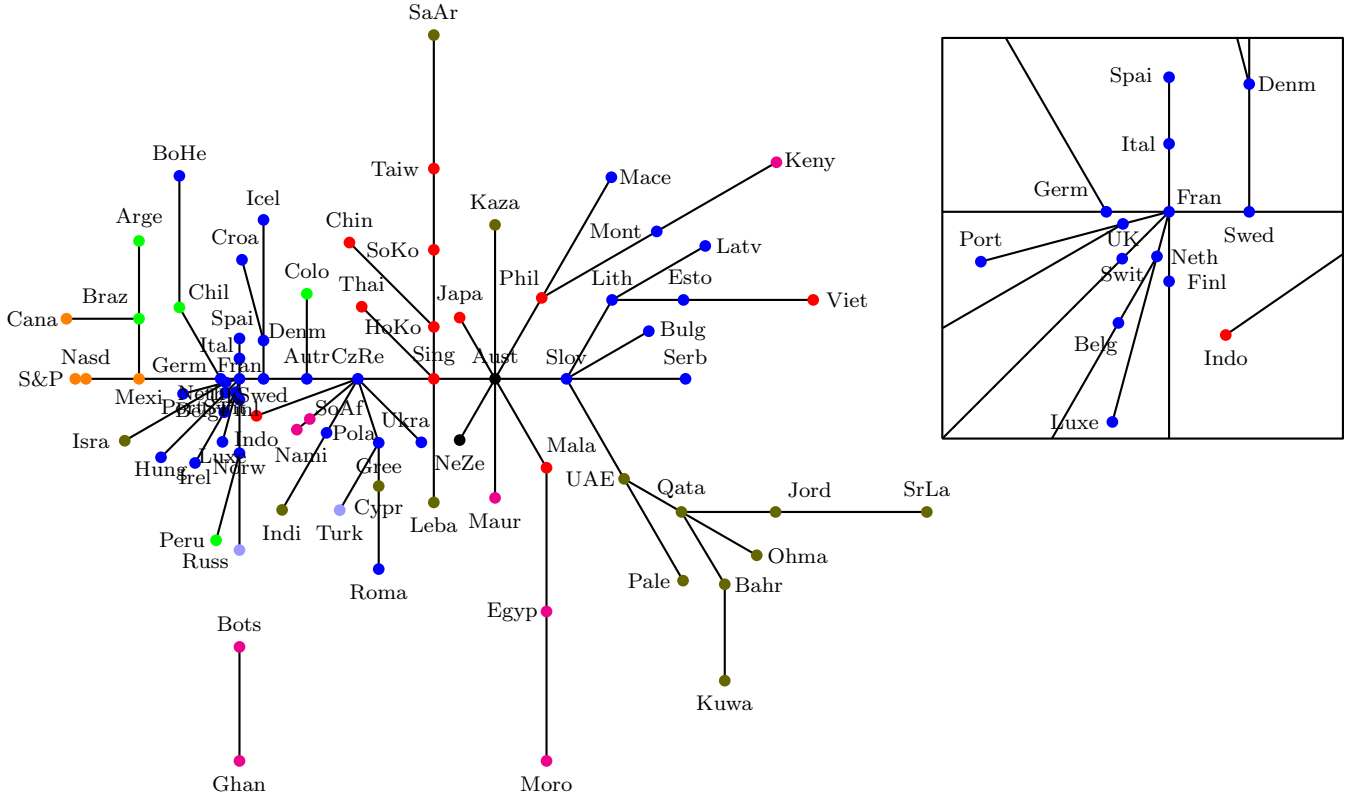


Figure 39: pruned minimum spanning tree for the second semester of 2008. The inset shows an amplification of the connections closest to France.

Note that the networks (many are not trees anymore) are much more compact, and most connections make sense if one considers other factors like geographic proximity, culture affinity, and economic ties between countries. Once again, one may notice that the networks shrink substantially during periods of international financial crises.

A last discussion shall be made in this section. The approach of choosing only strongly connected nodes, and also those weakly connected ones that endure in time, has produced graphs that are not substantially different from the ones we could have obtained by just pruning away every weakly interacting nodes. As checking for survivability is a time consuming task, most of the times there will be no real damage in pruning a minimum spanning tree using only the first criterium. Nevertheless, other types of networks, based solely on survivability, may also be built, with their own merits and maladies.

6 Centrality measures

A substantial part of network theory is devoted to the idea of the centrality of a node [50], or how influential the vertex is in the network. It is an important measurement which is handled in a number of different ways. In what follows, I perform an analysis of the centrality of vertices in the networks depicted by the minimum spanning trees according to three different definitions of centrality. I also do some analysis of the frequency distribution of each centrality measure in the network and which are the stocks that are more central according to each definition.

6.1 Node degree

The node degree of a node, or stock exchange index in this article, is the number of connections it has in the network. Most of the indices have low node degree, and some of them have a large node degree associated with them. The latter are called hubs, and are generally nodes that are more important in events that can change the network. Figures 40 to 43 show the evolution in the four different crisis periods we are studying of

the frequency distribution of the node degrees of the indices in the original minimum spanning trees (not the pruned ones). The graphs for the pruned networks are nearly identical to these ones. Note that the frequencies all drop exponentially, as it would be expected from scale free networks.

During the first crisis, 1986/1987, the indices with the largest node degree are Nasdaq and Japan (both with node degree 5), S&P, Canada, Netherlands, and Ireland (all with node degree 4); for 1997/1998, the indices with largest node degree are the Netherlands and Finland; for 200/2001, we have France, Netherlands, and Hong Kong occupying the first positions; for 2007/2008, this role is played by France, Germany, Singapore, and Australia.

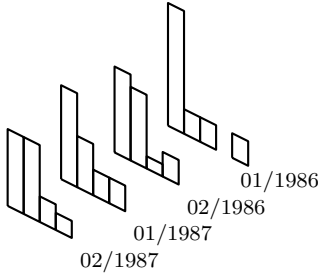


Figure 40: probability distributions of the node degree for the first and second semesters of 1986 and 1987.

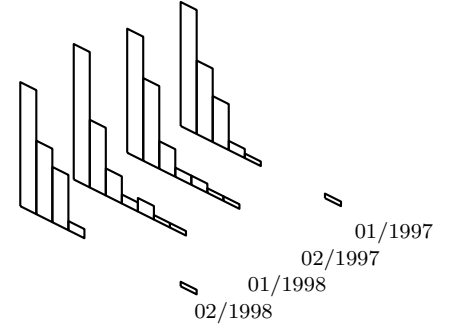


Figure 41: probability distributions of the node degree for the first and second semesters of 1997 and 1998.

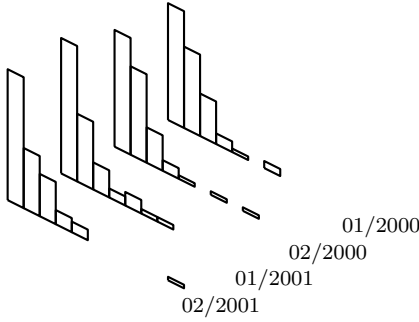


Figure 42: probability distributions of the node degree for the first and second semesters of 2000 and 2001.

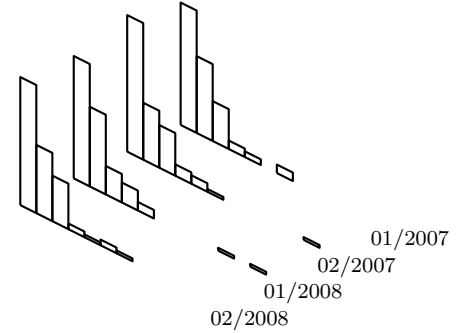


Figure 43: probability distributions of the node degree for the first and second semesters of 2007 and 2008.

6.2 Node strength

The strength of a node is the sum of the correlations of the node with all other nodes to which it is connected. If C is the matrix that stores the correlations between nodes that are linked in the minimum spanning tree, then the node strength is given by

$$N_s^k = \sum_{i=1}^n C_{ik} + \sum_{j=1}^n C_{kj} , \quad (5)$$

where C_{ij} is an element of matrix C .

Figures 44 to 47 show the evolution in the four different crisis periods of the frequency distribution of the node strengths of the indices in the original minimum spanning trees. The exponential drop seen for the node degrees distributions is now substituted by a strongly tilted distribution which is not exponentially decreasing.

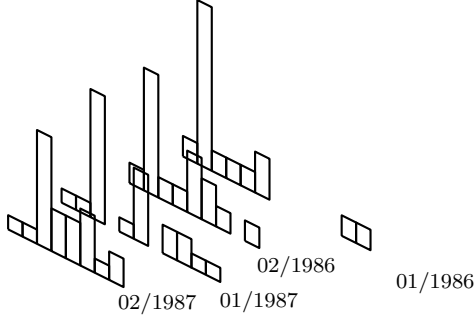


Figure 44: probability distributions of the node strength for the first and second semesters of 1986 and 1987.

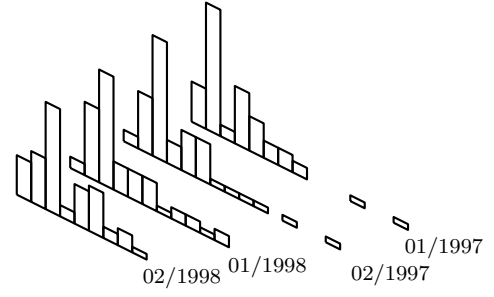


Figure 45: probability distributions of the node strength for the first and second semesters of 1997 and 1998.

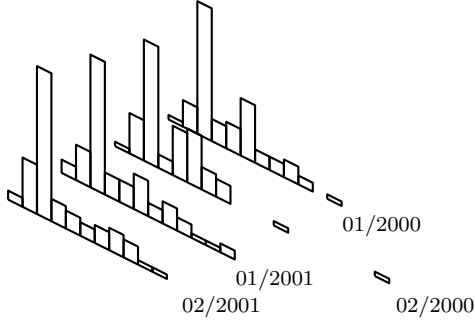


Figure 46: probability distributions of the node strength for the first and second semesters of 2000 and 2001.

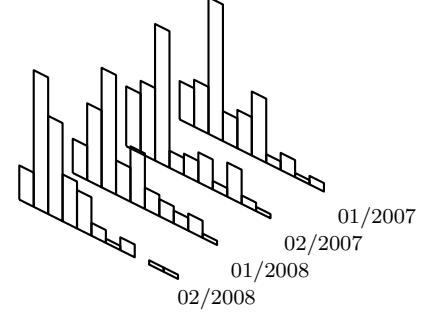


Figure 47: probability distributions of the node strength for the first and second semesters of 2007 and 2008.

6.3 Betweenness

The betweenness centrality measures how much a node lies on the shortest paths between other vertices. It is an important measure of how much a node is important as an intermediate between other nodes. It may be defined as

$$B_c^k = \sum_{i,j=1}^n \frac{n_{ij}^k}{m_{ij}}, \quad (6)$$

where n_{ij} is the number of shortest paths (geodesic paths) between nodes i and j that pass through node k and m_{ij} is the total number of shortest paths between nodes i and j . Our network is fully connected, so we need not worry about m_{ij} being zero. The frequency distribution of the betweenness of the indices in the original minimum spanning trees is shown in figures 48 to 51. The distributions drop substantially after the first interval.

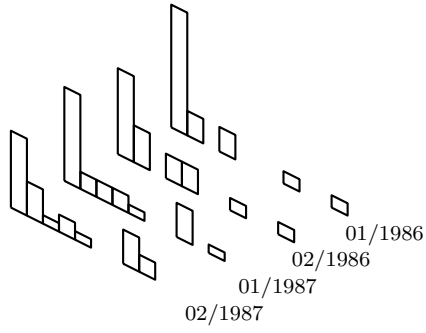


Figure 48: probability distributions of the normalized betweenness for the first and second semesters of 1986 and 1987.

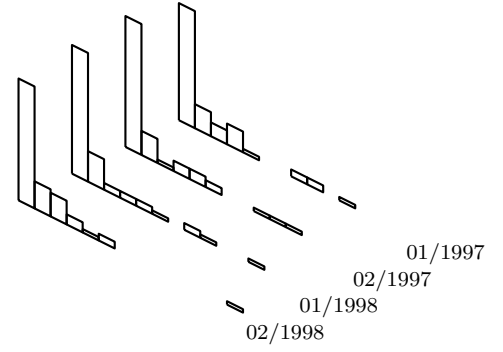


Figure 49: probability distributions of the normalized betweenness for the first and second semesters of 1997 and 1998.



Figure 50: probability distributions of the normalized betweenness for the first and second semesters of 2000 and 2001.

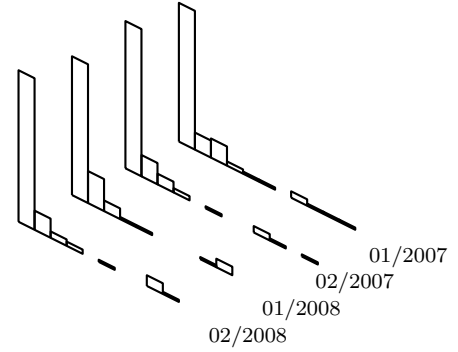


Figure 51: probability distributions of the normalized betweenness for the first and second semesters of 2007 and 2008.

Figure 52 shows the evolution in time of the three centrality measures whose frequency distributions have been just shown plus the Eigenvector centrality, which takes into account the degree of neighbouring nodes when calculating the importance of a node. In the figure, geographical regions are specified, and one may notice that Europe concentrates most of the centralities. The peaks are smeared out in the node strength representations, and are more evident in the betweenness centrality evolution. The countries of Oceania (Australia and New Zealand) show remarkable centralities, as do some key Pacific Asian countries. The USA is, surprisingly, not very central in these representations.

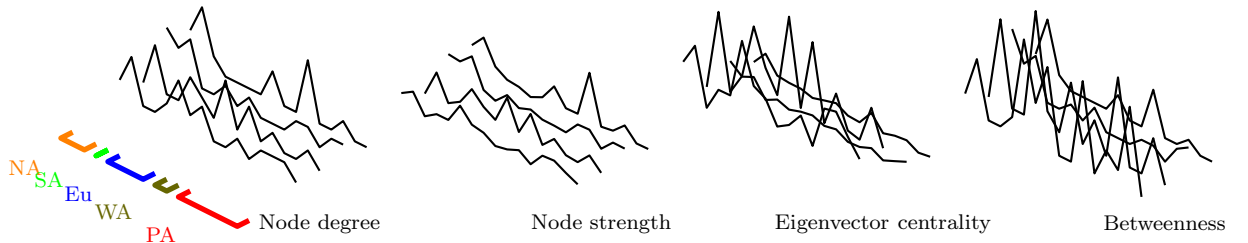


Figure 52a: first and second semesters of 1986 and 1987.

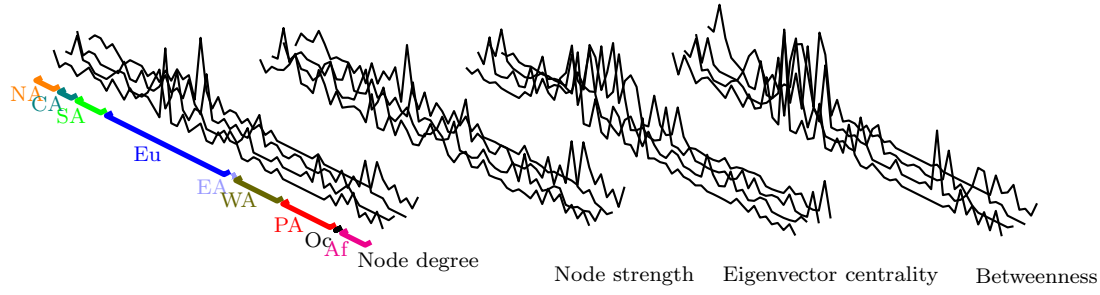


Figure 52b: first and second semesters of 1997 and 1998.

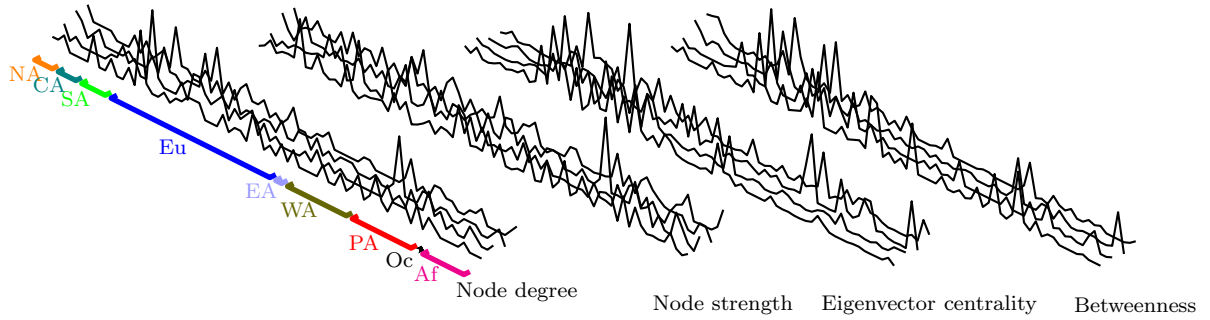


Figure 52c: first and second semesters of 2000 and 2001.

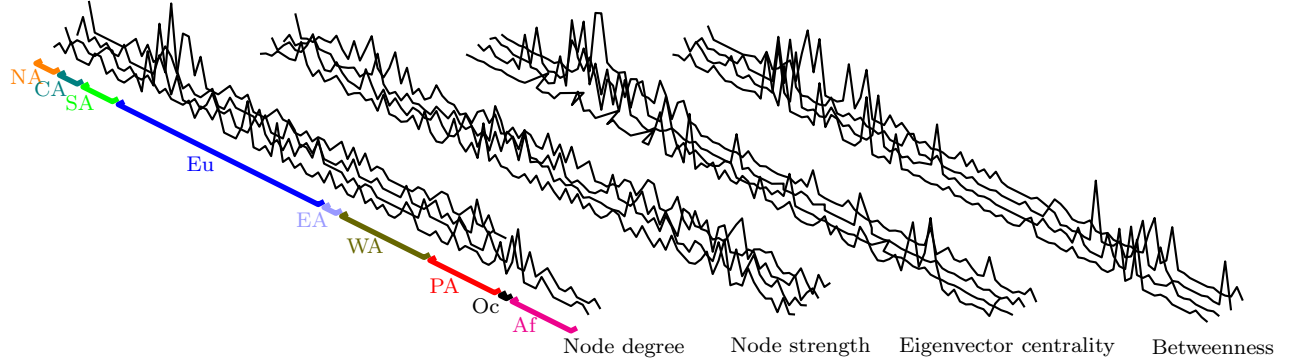


Figure 52d: first and second semesters of 2007 and 2008.

Figure 52: evolution of some centrality measures of stock exchange indices in time. From right to left, each graphs shows the evolution for the first and second semesters of, respectively, a) 1986 to 1987, b) 1997 to 1998, c) 2000 to 2001, and d) 2007 to 2008. The graphs also discriminate indices by geographic region: NA (orange) stands for North America, CA (darkgreen) for Central America and the Caribbean, SA (green) for South America, Eu (blue) for Europe, EA (light blue) for Eurasia, WA (brown) for Western Asia, PA (red) for Pacific Asia, Oc (black) for Oceania, and Af (magenta) for Africa.

7 Conclusion

Most of the previous works dealing with the minimum spanning tree representation of networks in finance only dealt with the most interacting nodes, what was done by choosing the most liquid stocks in a stock exchange, or the ones used in order to build an official index of the same stock exchange. So, poorly connected nodes never arised as an important issue, although some authors have taken this into account [47], using another

sort of pruning. In the approach of this article, I followed the path of limiting the connections to those that were strong, meaning they were related with distances below a threshold that was obtained by simulations with randomized data, and to those that were not strong, but that were very enduring in time. This made it possible to obtain diagrams that show sensible relations between the many indices that were studied. The graphs showed that the connections between stock market indices are strongly influenced by the geographic positions of the stock exchanges, as there are well defined clusters related with continents, reinforcing results previously obtained by regression methods [29]. They also show that cultural relations or international treaties have a strong effect in stock market indices correlations, as one can see by the strong connections between Israel and South Africa with Europe, and between Greece and Cyprus, as examples. The centrality of France has been detected by other authors [37], and the closeness between Russia and Turkey was detected in [51].

An unexpected result is the relative isolation of the USA stock markets, strongly linked with the other North and South American stock markets, but weakly linked with Europe. This is probably related with the fact that stock markets operate at different hours, so that the trading hours of most European markets intersect the opening hours of American ones only slightly, and Asian markets have null intersections with their American counterparts. How deep this affects the correlations between markets and how this discrepancy should be dealt with is a matter for further study. Preliminary studies made by the author show that there is no considerable change in the minimum spanning trees of world indices if one considers the stock market indices of countries that have null intersection with the NYSE lagged by one day, which means one compares the S&P 500 and Nasdaq indices with the next day indices of those markets. One remarkable exception is the index from New Zealand, which becomes much more connected with the USA indices when lagged by one day. The separation between Central Europe and North America persists even when one lags the whole of Europe by one day. The exception is the FTSE 100 from the UK, which in this case gets more connected with America than with Europe. A model developed in [52]-[53] for the analysis of intraday data, where comparing data taken at different times (or ticks) is an important issue, may also be applied to the lagging problem of international stock market indices.

The other factor to be observed is the nature of the centrality measures in a network. They are built in such a way that, if many nodes move together, then one or more of them are better representatives of the collective movement, like France with respect to the European market nodes. Centrality does not mean, necessarily, that the node is more influential. The S&P 500 (or the Dow Jones), for example, is watched by all stock market dealers in the world, although it is not very central in the minimum spanning trees. This is not due to the particularities of the minimum spanning tree representation, as other work [ref], which deals with asset graphs based on thresholds, show. What is needed is a measure of the impact a market has on others, a one-directional measure that takes into account the volume a stock market trades, and also the solidity and inertia of the market. Some measures which handle causality are partial correlation [54]-[57], Granger causality [58]-[60], and transfer of entropy [61]-[63], but none of them takes into account the size of a market.

Models of how a crisis may propagate on a network of stock markets have been built using epidemics models or econometric models, with some success [64]-[71]. Contagion models consider centrality or k-core number as key features in the propagation of crises.

This work contributes to a better understanding of the minimum spanning tree representation when dealing with weakly interacting indices, and offers a way to prune some of them away from the network. The process shows that weakly interacting nodes rarely survive for long, so that, in practice, it makes little difference if one prunes them away based only on a distance threshold obtained by randomizing the original data. It also shows how the world stock markets evolved in the past decades, mainly in the proximity of crises, and how stock markets became more integrated with time.

Further work, as cited above, shall deal with lag effects between markets, with the causality chain between them, and with the building of a better model that explains a little of the complex interactions between markets, how contagion between them happens, and maybe how to devise policies for cutting such contagions short.

Acknowledgements

The author acknowledges the support of this work by a grant from Insper, Instituto de Ensino e Pesquisa. I am also grateful to Siew Ann Cheong, Rudi Schaffer, Dror Kenett, and Sergio Giovanetti Lazzarini, for useful

discussions, and to the attenders and organizers of the Econophysics Colloquium 2010. This article was written using L^AT_EX, all figures were made using PSTricks, and the calculations were made using Matlab, Ucinet and Excel. All data are freely available upon request on leonidassj@insper.edu.br.

A Stock Market Indices

The next table (table 1) shows the stock market indices we used, their original countries, the symbols we used for them in the main text, and their codes in Bloomberg. In the tables, we use “SX” as short for “Stock Exchange”. Some of the indices changed names and/or method of calculation and are designated by the two names, prior to and after the changing date.

Index	Country	Symbol	Code in Bloomberg
North America			
S&P 500	United States of America	S&P	SPX
Nasdaq Composite	United States of America	Nasd	CCMP
S&P/TSX Composite	Canada	Cana	SPTSX
IPC	Mexico	Mexi	MEXBOL
Central America			
Bolsa de Panama General	Panama	Pana	BVPSBVPS
BCT Corp Costa Rica	Costa Rica	CoRi	CRSMBCT
Caribbean			
Jamaica SX Market Index	Jamaica	Jama	JMSMX
British overseas territories			
Bermuda SX Index	Bermuda	Berm	BSX
South America			
Ibovespa	Brazil	Braz	IBOV
Merval	Argentina	Arge	MERVAL
IPSA	Chile	Chil	IPSA
IGBC	Colombia	Colo	IGBC
IBC	Venezuela	Vene	IBVC
IGBVL	Peru	Peru	IGBVL
Western and Central Europe			
FTSE 100	United Kingdom	UK	UKX
ISEQ	Ireland	Irel	ISEQ
CAC 40	France	Fran	CAC
DAX	Germany	Germ	DAX
SMI	Switzerland	Swit	SMI
ATX	Austria	Autr	ATX
FTSE MIB or MIB-30	Italy	Ital	FTSEMIB
Malta SX Index	Malta	Malt	MALTEX
BEL 20	Belgium	Belg	BEL20
AEX	Netherlands	Neth	AEX
Luxembourg LuxX	Luxembourg	Luxe	LUXXX
OMX Stockholm 30	Sweden	Swed	OMX
OMX Copenhagen 20	Denmark	Denm	KFX
OMX Helsinki	Finland	Finl	HEX
OBX	Norway	Norw	OBX
OMX Iceland All-Share Index	Iceland	Icel	ICEXI
IBEX 35	Spain	Spai	IBEX
PSI 20	Portugal	Port	PSI20
Athens SX General Index	Greece	Gree	ASE

Index	Country	Symbol	Code in Bloomberg
Eastern Europe			
PX or PX50	Czech Republic	CzRe	PX
SAX	Slovakia	Slok	SKSM
Budapest SX Index	Hungary	Hung	BUX
BELEX 15	Serbia	Serb	BELEX15
CROBEX	Croatia	Croa	CRO
SBI TOP	Slovenia	Slov	SBITOP
SASE 10	Bosnia and Herzegovina	BoHe	SASX10
MOSTE	Montenegro	Mont	MOSTE
MBI 10	Macedonia	Mace	MBI
WIG	Poland	Pola	WIG
BET	Romania	Roma	BET
SOFIX	Bulgaria	Bulg	SOFIX
OMXT	Estonia	Esto	TALSE
OMXR	Latvia	Latv	RIGSE
OMXV	Lithuania	Lith	VILSE
PFTS	Ukraine	Ukra	PFTS
Eurasia			
MICEX	Russia	Russ	INDEXCF
ISE National 100	Turkey	Turk	XU100
Western and Central Asia			
KASE	Kazakhstan	Kaza	KZKAK
CSE	Cyprus	Cypr	CYSMMAPA
Tel Aviv 25	Israel	Isra	TA-25
Al Quds	Palestine	Pale	PASISI
BLOM	Lebanon	Leba	BLOM
ASE General Index	Jordan	Jord	JOSMGNFF
TASI	Saudi Arabia	SaAr	SASEIDX
Kuwait SE Weighted Index	Kuwait	Kuwa	SECTMIND
Bahrain All Share Index	Bahrein	Bahr	BHSEASI
QE or DSM 20	Qatar	Qata	DSM
ADX General Index	United Arab Emirates	UAE	ADSMI
MSM 30	Ohman	Ohma	MSM30
South Asia			
Karachi 100	Pakistan	Paki	KSE100
SENSEX 30	India	Indi	SENSEX
Colombo All-Share Index	Sri Lanka	SrLa	CSEALL
DSE General Index	Bangladesh	Bang	DHAKA
Oceania			
S&P/ASX 200	Australia	Aust	AS51
NZX 50	New Zealand	NeZe	NZSE50FG
Northern Africa			
CFG 25	Morocco	Moro	MCSINDEX
TUNINDEX	Tunisia	Tuni	TUSISE
EGX 30	Egypt	Egyp	CASE

Index	Country	Symbol	Code in Bloomberg
Asia-Pacific			
Nikkei 25	Japan	Japa	NKY
Hang Seng	Hong Kong	HoKo	HSI
Shangai SE Composite	China	Chin	SHCOMP
MSE TOP 20	Mongolia	Mong	MSETOP
TAIEX	Taiwan	Taiw	TWSE
KOSPI	South Korea	SoKo	KOSPI
SET	Thailand	Thai	SET
VN-Index	Vietnam	Viet	VNINDEX
KLCI	Malaysia	Mala	FBMKLCI
Straits Times	Singapore	Sing	FSSTI
Jakarta Composite Index	Indonesia	Indo	JCI
PSEi	Philippines	Phil	PCOMP
Central and Southern Africa			
Ghana All Share Index	Ghana	Ghan	GGSEGSE
Nigeria SX All Share Index	Nigeria	Nige	NGSEINDX
NSE 20	Kenya	Keny	KNSMIDX
DSEI	Tanzania	Tanz	DARSDSEI
FTSE/Namibia Overall	Namibia	Nami	FTN098
Gaborone	Botswana	Bots	BGSMDC
FTSE/JSE Africa All Share	South Africa	SoAf	JALSH
SEMDEX	Mauritius	Maur	SEMDEX

Table 1: names, codes, and abbreviations of the stock market indices used in this article.

B Minimum spanning trees in three dimensions

In what follows, I show the three dimensional pictures of the minimum spanning tree, both original and pruned, in three dimensions, using an algorithm that represents as best as possible the real distances between indices. Note that some indices that seem to be far apart in the two dimensional view of the minimum spanning trees are in fact close together, like some african indices displayed in the next figures.

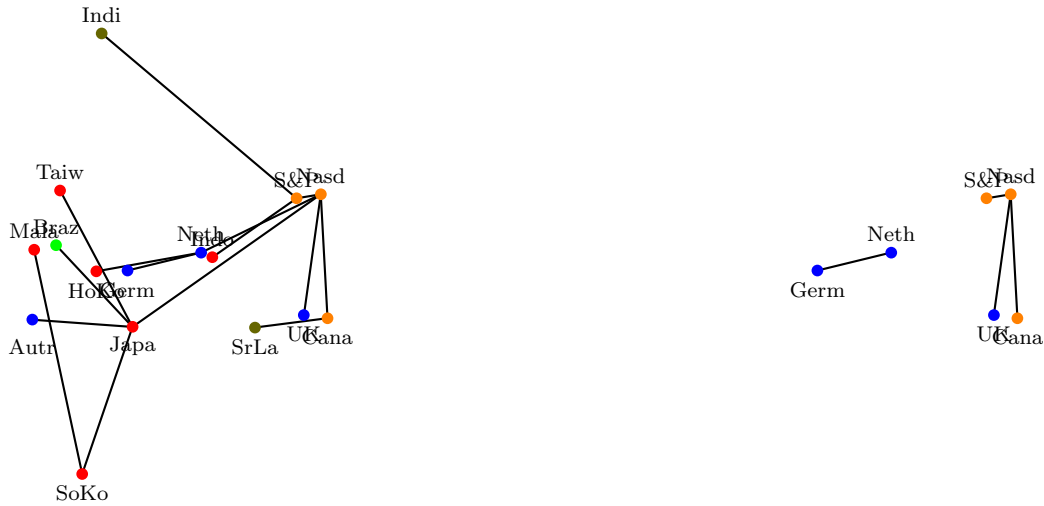


Figure 53: three dimensional view of the minimum spanning tree for the first semester of 1986. The second figure is the pruned spanning tree.

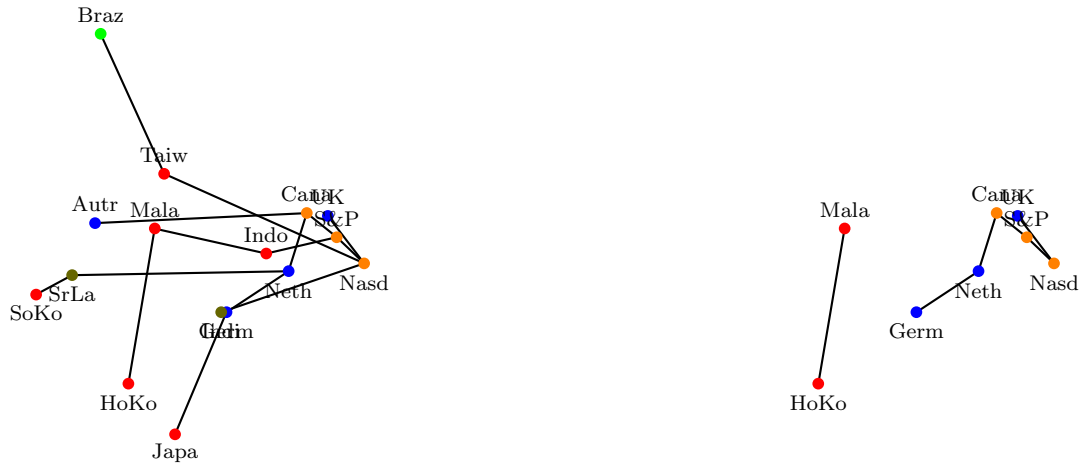


Figure 54: three dimensional view of the minimum spanning tree for the second semester of 1986. The second figure is the pruned spanning tree.

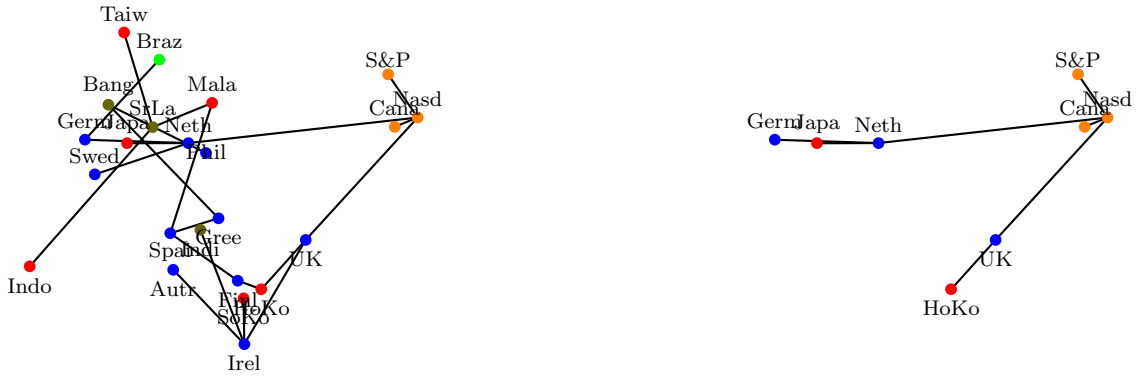


Figure 55: three dimensional view of the minimum spanning tree for the first semester of 1987. The second figure is the pruned spanning tree.

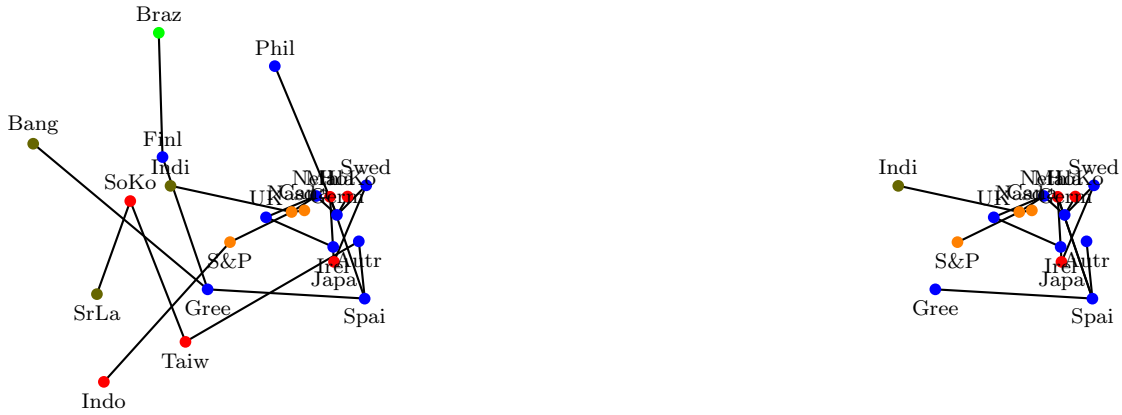


Figure 56: three dimensional view of the minimum spanning tree for the second semester of 1987. The second figure is the pruned spanning tree.

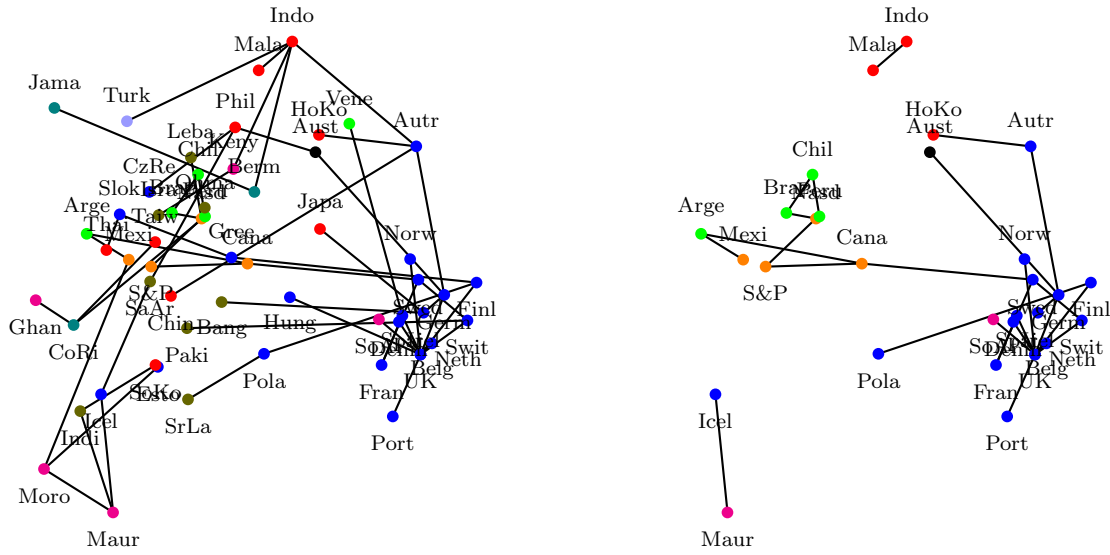


Figure 57: three dimensional view of the minimum spanning tree for the first semester of 1997. The second figure is the pruned spanning tree.

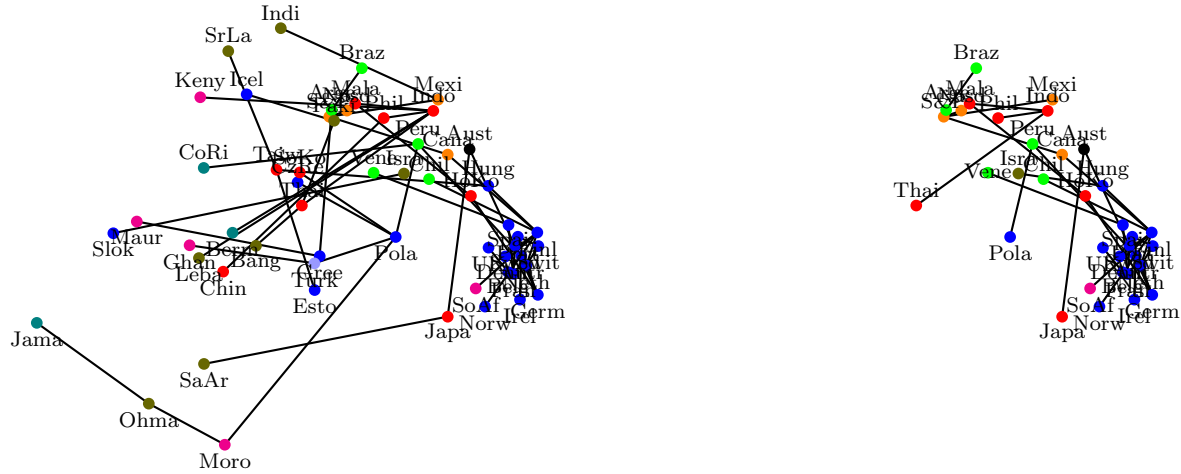


Figure 58: three dimensional view of the minimum spanning tree for the second semester of 1997. The second figure is the pruned spanning tree.

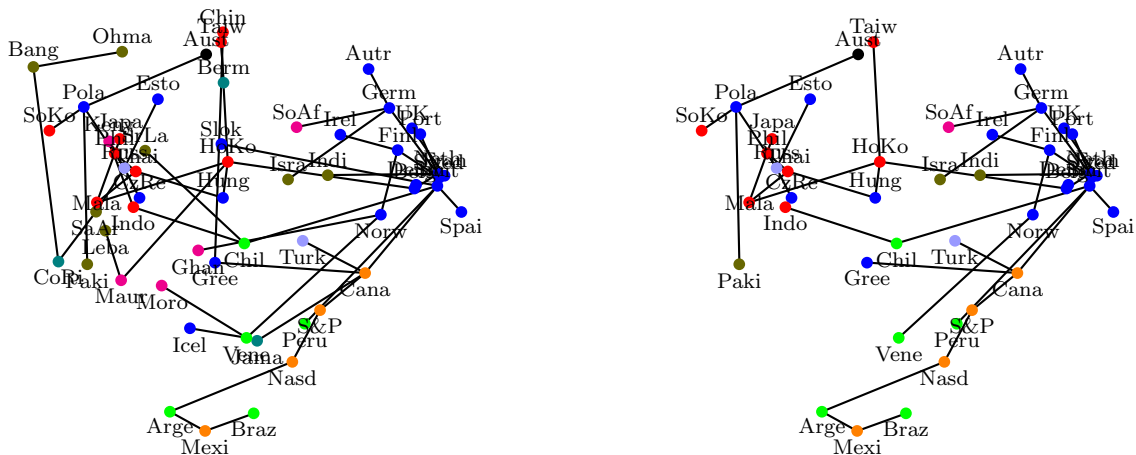


Figure 59: three dimensional view of the minimum spanning tree for the first semester of 1998. The second figure is the pruned spanning tree.

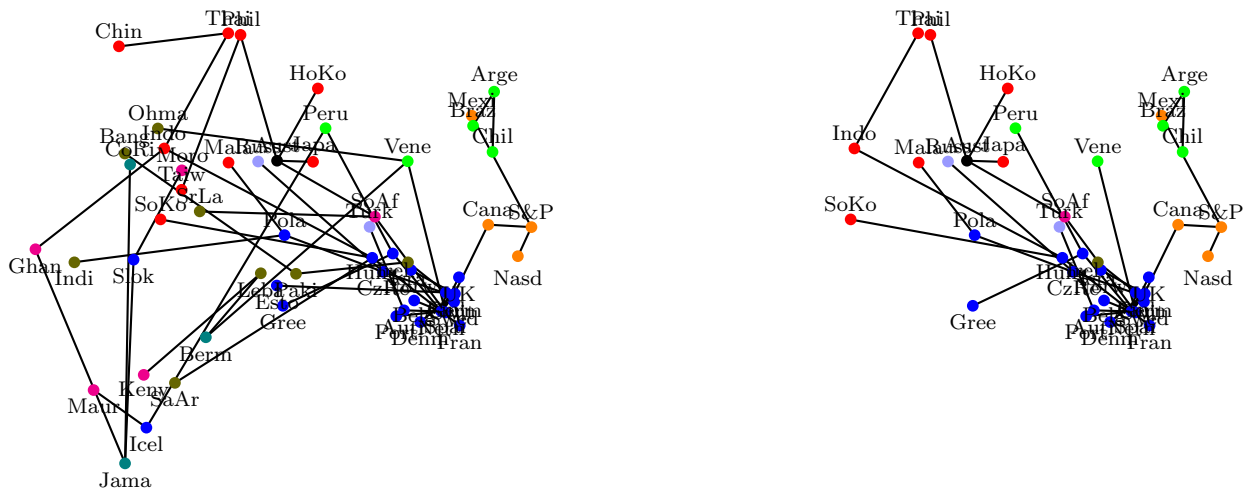


Figure 60: three dimensional view of the minimum spanning tree for the second semester of 1998. The second figure is the pruned spanning tree.

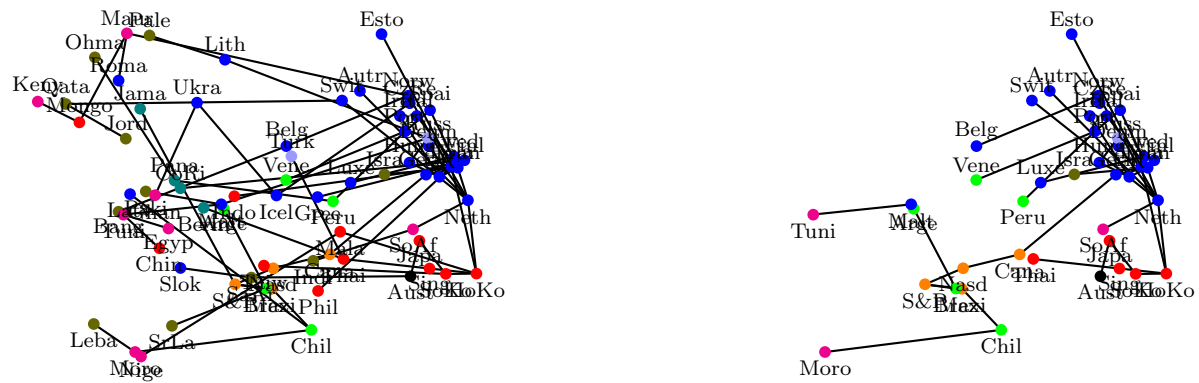


Figure 61: three dimensional view of the minimum spanning tree for the first semester of 2000. The second figure is the pruned spanning tree.

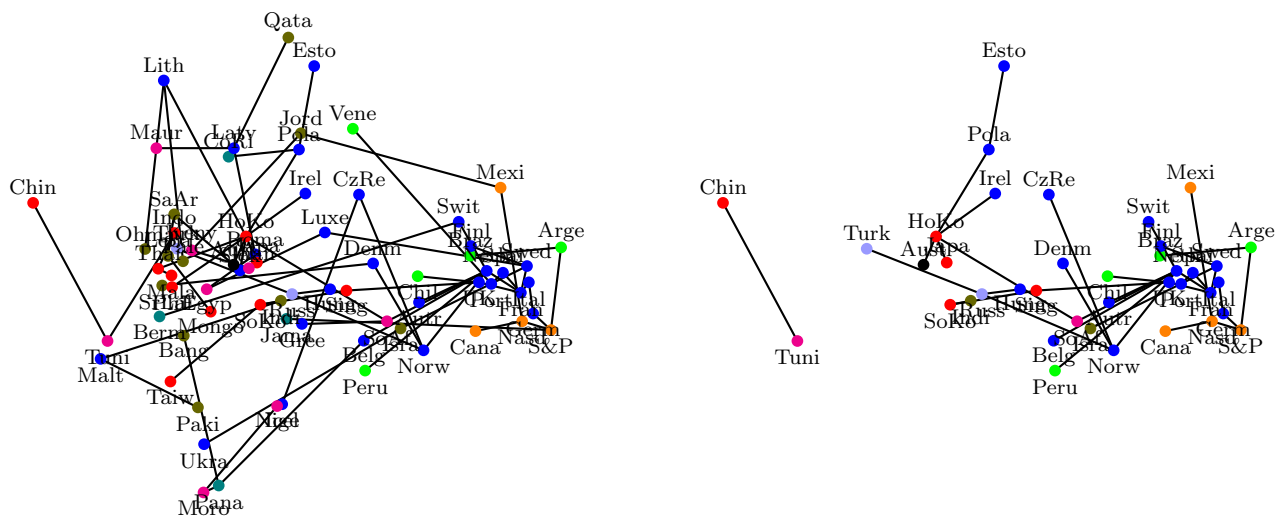


Figure 62: three dimensional view of the minimum spanning tree for the second semester of 2000. The second figure is the pruned spanning tree.

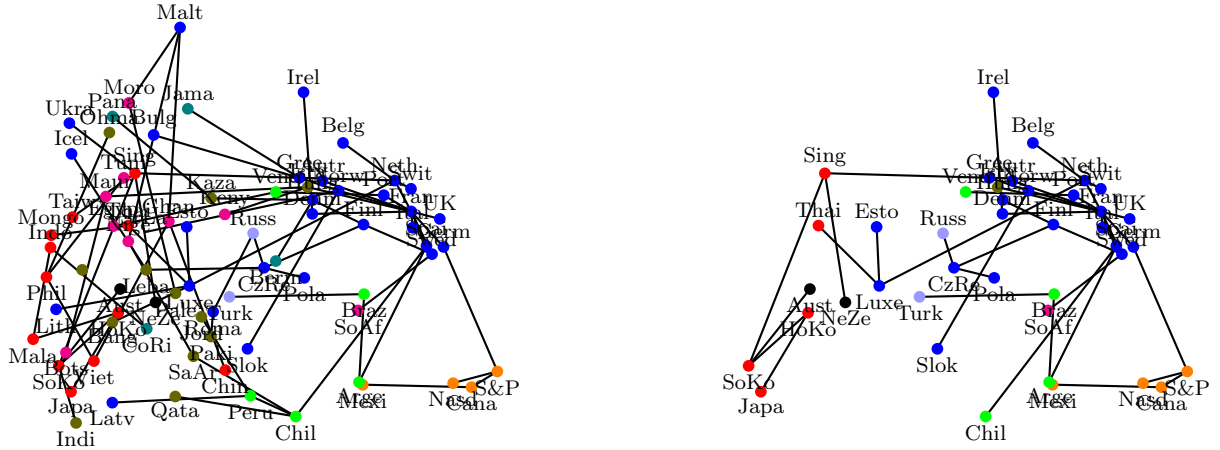


Figure 63: three dimensional view of the minimum spanning tree for the first semester of 2001. The second figure is the pruned spanning tree.

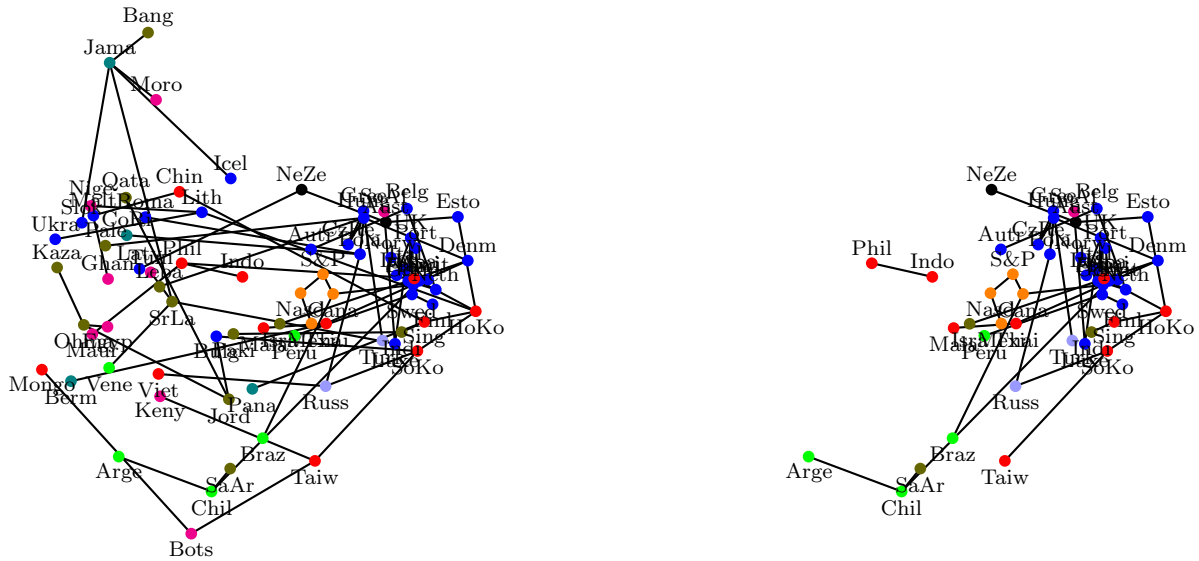


Figure 64: three dimensional view of the minimum spanning tree for the second semester of 2001. The second figure is the pruned spanning tree.

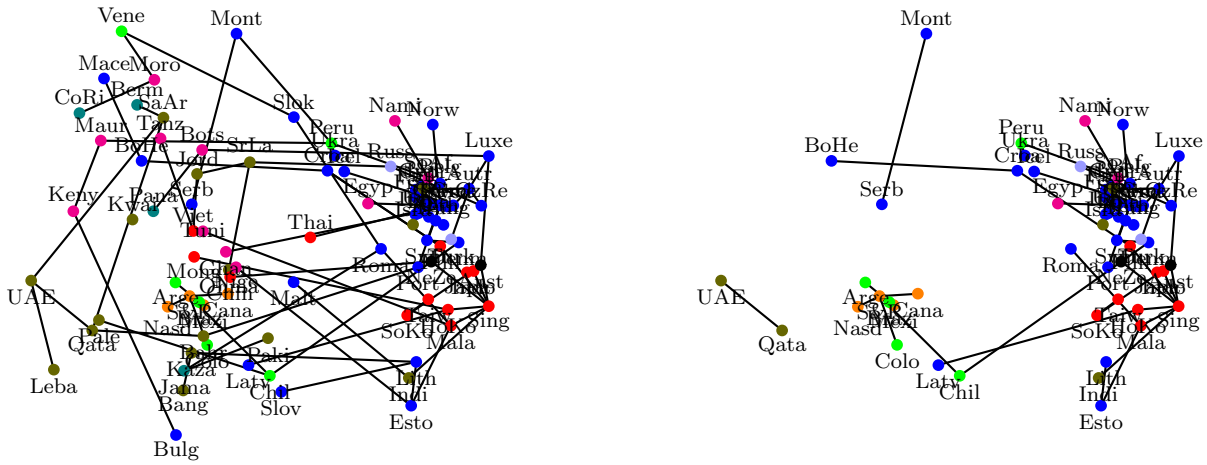


Figure 65: three dimensional view of the minimum spanning tree for the first semester of 2007. The second figure is the pruned spanning tree.

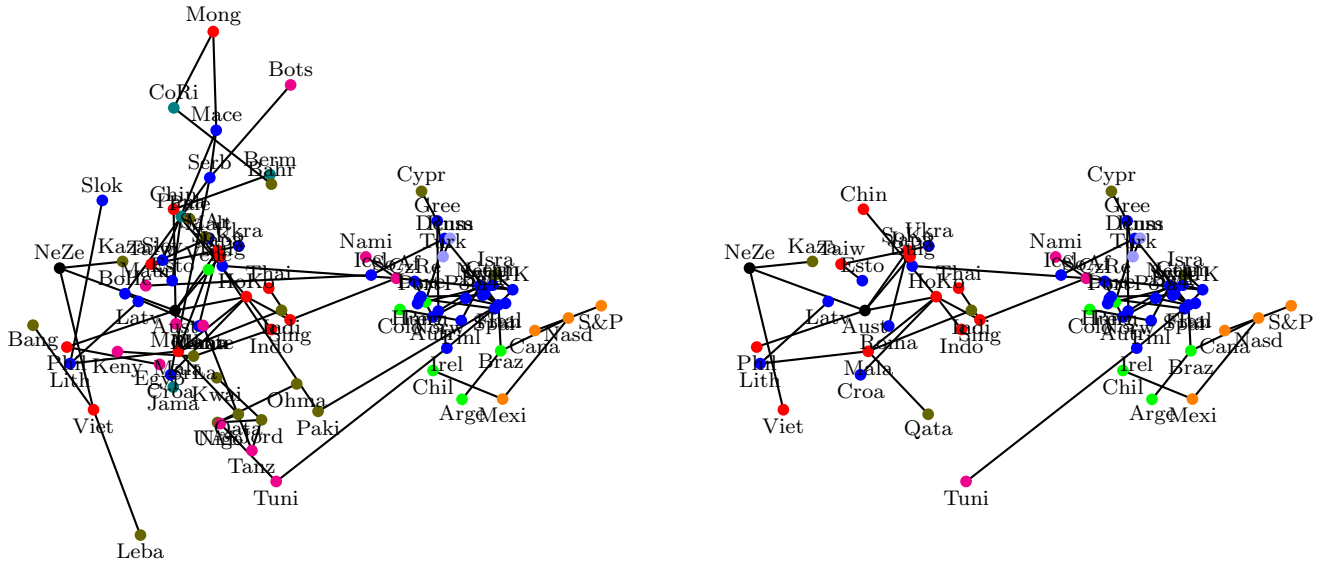


Figure 66: three dimensional view of the minimum spanning tree for the second semester of 2007. The second figure is the pruned spanning tree.

- [7] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertész, and A. Kanto, *Asset trees and asset graphs in financial markets*, Phys. Scripta T **106** (2003) 48-54.
- [8] J.-P. Onnela, K. Kaski, and J. Kertész, *Clustering and information in correlation based financial networks*, Eur. Phys. J. B **38** (2004) 353-362.
- [9] G. Bonanno, G. Caldarelli, F. Lillo, S. Micciché, N. Vandewalle, and R.N. Mantegna, *Networks of equities in financial markets*, The European Phys. J. B **38**, (2004) 363–371.
- [10] C. Coronello, M. Tumminello, F. Lillo, S. Micchichè, and R.N. Mantegna, *Sector identification in a set of stock return time series traded at the London Stock Exchange*, Acta Phys. Pol. B **36** (2005) 2653-2679.
- [11] T. Aste and T. Di Matteo, *Correlation filtering in financial time series*, Noise and Fluctuations in Econophysics and Finance, Proceedings of the SPIE 5848 (2005) 100-109.
- [12] S. Sinha and R.K. Pan, *Uncovering the internal structure of the Indian financial market: cross-correlation behavior in the NSE*, in “Econophysics of markets and business networks”, Springer (2007) 215-226.
- [13] C. Coronello, M. Tumminello, F. Lillo, S. Micchichè, and R.N. Mantegna, *Economic sector identification in a set of stocks traded at the New York Stock Exchange: a comparative analysis*, Proceedings of the SPIE, vol. 6601, 66010T (2007).
- [14] M. Tumminello, T. Di Matteo, T. Aste, and R.N. Mantegna, *Correlation based networks of equity returns sampled at different time horizons*, Eur. Phys. J. B **55** (2007) 209–217.
- [15] C. Borghesi, M. Marsili, and S. Micciché, *Emergence of time-horizon invariant correlation structure in financial returns by subtraction of the market mode*, Phys. Rev. E **76** (2007), 026104.
- [16] V. Kulkarni and N. Deo, *Correlation and volatility of an Indian stock market: a random matrix approach*, Eur. Phys. J. B **60** (2007) 101-109.
- [17] J.G. Brida and W.A. Risso, *Multidimensional minimal spanning tree: the Dow Jones case*, Physica A **387** (2008) 5205-5210.
- [18] Y.W. Goo, T.W. Lian, W.G. Ong, W.T. Choi, and S.A. Cheong, *Financial atoms and molecules*, preprint (2009) arXiv:0903.2009v1.
- [19] M. Tumminello, F. Lillo, and R.N. Mantegna, *Correlation, hierarchies, and networks in financial markets*, Journal of Economic Behavior & Organizations (2010) Article in press.
- [20] S. Drożdż, J. Kwapien, and J. Speth, *Coherent patterns in nuclei and in financial markets*, AIP Conf. Proc. **1261** (2010), 256-264.
- [21] Y. Zhanga, G.H.T. Leea, J.C. Wonga, J.L. Kokb, M. Prustyb, and S.A. Cheong, *Will the US Economy Recover in 2010? A Minimal Spanning Tree Study*, (2010) Physica A **390** (2011) 2020-2050.
- [22] S.A. Cheong, R.P. Fornia, G.H.T. Lee, J.L. Kok, W.S. Yim, D.Y. Xu, Y. Zhang, *The Japanese Economy in Crises: A Time Series Segmentation Study*, Economics **2011-24** (2011)
- [23] N.F. Johnson, M. McDonald, O. Suleman, S. Williams, and S. Howison, *What shakes the FX tree? Understanding currency dominance, dependence, and dynamics*, Noise and Fluctuations in Econophysics and Finance. Edited by D. Abbott, J.-P. Bouchaud, , X. Gabaix, and J.L. McCauley, Proceedings of the SPIE, Volume 5848, (2005) 86-99.
- [24] M.J. Taylor, L.C. Rose, and B.J. Moyle, *Topology of foreign exchange markets using hierarchical structure methods*, Physica A **382**, (2007) 199-208.
- [25] J. Kwapien, S. Gworek, S. Drożdż, and A. Górski, *Analysis of a network structure of the foreign currency exchange market*, J. of Economic Interaction and Coordination **4** (2009), 55-72.

- [26] J. Kwapień, S. Gworek, and S. Drożdż, *Structure and evolution of the foreign exchange networks*, Acta Physica Polonica B **40** (2009), 175-194.
- [27] D.J. Fenn, M.A. Porter, P.J. Mucha, M. McDonald, S. Williams, N.F. Johnson, and N.S. Jones, *Dynamical clustering of exchange rates*, preprint (2010) arXiv:0905.4912v2.
- [28] M. Keskin, B. Deviren, and Y. Kocalkaplan, *Topology of the correlation networks among major currencies using hierarchical structure methods*, Physica A **390** (2011), 719-730.
- [29] G. Fagiolo, *The international-trade network: gravity equations and topological properties*, J. of Economic Interaction and Coordination **5** (2010), 1-25.
- [30] J. He and M.W. Deem, *Structure response in the world trade network*, Phys. Rev. Lett. **105** (2010), 198701-1 to 198701-4.
- [31] L. De Benedictis and L. Tajoli, *The World Trade Network* (2009), forthcoming in The World Economy.
- [32] K.M. Lee, J.-S. Yang, J. Lee, K.-I. Goh, and I.- M. Kim, *Impact of the topology of global macroeconomic network on the spreading of economic crises*, (2011)
PLoS ONE 6(3): e18443.doi:10.1371/journal.pone.0018443
- [33] P. Sieczka and J.A. Hołyst, *Correlations in commodity markets*, Physica A **388** (2009), 1621-1630.
- [34] M. Ausloos and R. Lambiotte, *Clusters or networks of economies? A macroeconomy study through gross domestic product*, Physica A **382** (2007) 16-21.
- [35] S. Vitali, J.B. Glotfelder, and S. Battiston, *The network of global corporate control*, (2011) arXiv:1107.5728v1.
- [36] D. Harmon, B. Stacey, Y. Bar-Yam, and Y. Bar-Yam, *Networks of Economic Market Interdependence and Systemic Risk*, (2011) arXiv: 1011.3707v2.
- [37] R. Coelho, C.G. Gilmore, B. Lucey, P. Richmond, and S. Hutzler, *The evolution of interdependence in world equity markets - evidence from minimum spanning trees*, Physica A **376** (2007) 455-466.
- [38] C. Eom, G. Oh, W.-S. Jung, H. Jeong, and S. Kim, *Topological properties of stock networks based on minimal spanning tree and random matrix theory in financial time series*, Physica A **388** (2009), 900-906.
- [39] M. Eryiğit and R. Eryiğit, *Network structure of cross-correlations among the world market indices*, Physica A **388** (2009), 3551-3562.
- [40] D.-M. Song, M. Tumminello, W-X Zhou, and R.N. Mantegna, *Evolution of worldwide stock markets, correlation structure and correlation based graphs*, Phys. Rev. E **84**, (2011) 026108.
- [41] M. Fischetti, H. Hamacher, K. Jörnsten, and F. Maffoli, *Weighted k-cardinality trees: Complexity and polyhedral structure*, Networks **24** (1994), 1298-1309.
- [42] M.X. Goemans, *Minimum bounded degree spanning trees*, preprint.
- [43] M. Ghodsi, H. Mahini, K. Mirjalali, S.O. Gharan, A.S. Sayedi R., and M. Zadimoghaddam, *Spanning trees with minimum weighted degrees*, Information Processing Letters **104**, (2007) 113-116.
- [44] L. Sandoval Jr. and I. De P. Franca, *Correlation of financial markets in times of crisis*, (2011) Physica A (in press).
- [45] L. Kullmann, J. Kertész, and K. Kaski, *Time dependent cross correlations between different stock returns: A directed network of influence*, Phys. Rev. E **66**, (2002) 026125.
- [46] J.-P. Onnela, A. Chakraborti, K. Kaski, and J. Kertesz, *Dynamic asset trees and Black Monday*,

- [47] D. Lautier and F. Raynaud, *Systemic risk in derivative markets: a graph-theory analysis*, (July 05, 2011). Available at SSRN: <http://ssrn.com/abstract=1579629>.
- [48] M. McDonald, O. Suleman, S. Williams, S. Howison, and N.F. Johnson, *Detecting a currency's dominance or dependence using foreign exchange network trees*, Phys. Rev. E **72**, (2005) 046106.
- [49] A. Buda, *Lifetime of correlations between stocks on established and emerging markets*, Life-time of correlation and its applications (volume 1), (2010) Wydawnictwo Niezależne, Wrocław, available at <http://science24.com/paper/17818>. Physica A **324**, (2003) 247-252.
- [50] M.E.J. Newman, *Networks, and introduction*, (2010) Oxford University Press.
- [51] I. Aktar, *Is there any comovement between stock markets of Turkey, Russia and Hungary?*, Int. Res. J. of Fin. and Economics **26**, (2009) 192-200.
- [52] R. Schäfer, M. Sjöllin, M. Wolanski, and T. Guhr, *Credit Risk - A structural model with jumps and correlations*, Physica A **383**, (2007) 533-569.
- [53] R. Schäfer, M. Sjöllin and T. Guhr, *Local normalization: Uncovering correlations in non-stationary financial time series*, Physica A **389**, (2010) 3856-3865.
- [54] D.Y. Kenett, Y. Shapira, and E. Ben-Jacob, *RMT assessments of the market latent information embedded in the stock's raw, normalized, and partial correlations*, J. Prob. Stat. **2009**, (2009) 1-13.
- [55] D.Y. Kenett, M. Tumminello, A. Madi, G. Gur-Gershgoren, R. Mategna, and E. Ben-Jacob, *Dominating clasp of the financial sector revealed by partial correlation analysis of the stock market*, Plos One **5**, (2010).
- [56] D.Y. Kenett, Y. Shapira, A. Madi, S. Bransburg-Zabary, G. Gur-Gershgoren, and E. Ben-Jacob, *Dynamics of stock market correlations*, AUCO Czech Economic Review **4**, (2010) 330-340.
- [57] Y. Shapira, D.Y. Kenett, and E. Ben-Jacob, *The index cohesive effect on stock market correlations*, Eur. Phys. J. B **72**, (2009) 657-669.
- [58] D.A. Bessler and J. Yang, *The structure of interdependence in international stock markets*, J. of Int. Money and Finance **22**, (2003) 261-287.
- [59] A. Allali, A. Oueslati, and A. Trabelsi, *Detection of Information Flow in Major International Financial Markets by Interactivity Network Analysis*, Asia-Pacific Financial Markets **18**, (2011) 319-344.
- [60] A. Abdelwahab, O. Amor, and T. Abdelwahed, *The analysis of the interdependence structure in international financial markets by graphical models*, International Research Journal of Finance and Economics **15** (2008) 291-306.
- [61] T. Schreiber, *Measuring information transport*, Phys. Rev. Lett. **85**, (2000) 461-464.
- [62] O. Kwon and J-S Yang, *Information flow between stock indices*, Europh. Lett. **82**, (2008) 68003.
- [63] O. Kwon and J-S Yang, *Information flow between composite stock index and individual stocks*, Physica A **387**, (2008) 2851-2856.
- [64] R. Rigobon, *On the measurement of the international propagation of shocks: is the transmission stable?*, J. Int. Economics **61**, (2003) 261-283.
- [65] I. Simonsen, *Diffusion and networks: A powerful combination!*, Physica A **357**, (2005) 317-330.
- [66] R. Smith, *The Spread of the Credit Crisis: View from a Stock Correlation Network*, J. of the Korean Phys. Society **54**, (2009) 2460-2463.
- [67] M. Kitsak, L.K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H.E. Stanley, and H.A. Makse, *Identifying influential spreaders in complex networks*, Nature Physics **6**, (2010) 888-893.

- [68] T. Komatsu and A. Nomatame, *Diffusion in optimized networks*, Int. J. of Bio-Inspired Computation (2011) (in press).
- [69] Z. Ahmedov and D.A. Bessler, *Structure of interdependencies among international stock markets and contagion patterns of 2008 global financial crisis*, (2011) preprint available at <http://purl.umn.edu/98858>.
- [70] M. Beine, G. Capelle-Blancard, and H. Raymond, *International nonlinear causality between stock markets*, The Eur. J. of Finance **14**, (2008) 663-686.
- [71] P. Gai and S. Kapadia, *Contagion in financial networks*, Working paper 383 (2010), Bank of England.